2010

Measuring the Discriminative Capability of Metrology in Human Recognition

Rakesh Sepuri
West Virginia University

Measuring the Discriminative Capability of Metrology in Human Recognition

by

Rakesh Sepuri

Thesis submitted to the College of Engineering and Mineral Resources at West Virginia University

in partial fulfillment of the requirements for the degree of

Master of Science
in
Electrical Engineering (Communications & Signal Processing)

Approved by

Dr. Donald Adjeroh, Committee Chairperson
Dr. Arun Ross
Dr. Xin Li

Lane Department of Computer Science and Electrical Engineering

Morgantown, West Virginia
2010

Keywords: Metrology, CAESAR dataset, biometric system, feature extraction, soft biometrics, statistical analysis, performance measurement

ABSTRACT

Measuring the Discriminative Capability of Metrology in Human Recognition

By Rakesh Sepuri

Biometrics is now a more established and sophisticated field. It studies the use of physiological characteristics or behavioral traits to identify and recognize a person automatically. It facilitates theft control and increased security.

Metrology has been one of the well-studied topics in computer vision. Absolute measurement values of humans can be obtained from a fully calibrated camera. These measurements are stored as a database and studied in detail to assess their significance as a biometric. In this thesis we want to assess the performance of human body measurements as a soft biometric. Every human has distinct biometric characters. They can be classified using biometric measurements.

Here the performance of Biometric systems is measured empirically without explicitly measuring the available information. We make use of soft biometric traits like height, weight, gender, age to measure the discriminative capability of metrology in human recognition. Analysis of human body measurements can be applied in various domains like video surveillance, video retrieval, human-computer interaction systems, and medical diagnosis. We establish the performance of human metrology in distinguishing between humans using a database of such measurements. We characterize the performance using measures such as distance plots, precision and recall, genuine acceptance rate (GAR), and false acceptance rate (FAR).

DEDICATION

I dedicate this thesis to my parents and brother

Ravi Kumar Sepuri
Susheela Sepuri
Sumesh Sepuri

ACKNOWLEDGMENTS

This work would not have been complete without the guidance and help of several individuals who directly or indirectly have contributed and extended their valuable assistance in the completion of this thesis.

First and foremost, I would like to take the opportunity to thank my advisor and committee head Dr. Donald Adjeroh for his valuable guidance and advice. He supported me throughout my work with his timely suggestions and patience. He inspired me greatly in working through this project. His ideas are innovative, and his problem solving skills are commendable. He has always motivated me to work with zeal in my research work. I feel highly honored to work under him and making my Masters program a memorable experience.

I would also like to thank Dr. Arun Ross under whom I learnt the course *Pattern Recognition*, which partly forms a basis for this work. His lectures were very helpful in doing my ground work. His command over the subject of pattern recognition is supreme. I am very grateful to Dr. Xin Li too for being on my committee.

Also, I would like to thank my family members, friends and well wishers for their understanding and support. They were always cordial and encouraged me in my work, providing a conducive environment to carry out this work.

Last but not the least, I give thanks to the Almighty Lord for answering my prayers and giving me strength in difficult times.

TABLE OF CONTENTS

LIST OF FIGURES

LIST OF TABLES

Chapter 1: Introduction

1.1 Introduction

We see many kinds of images in our daily life, surrounding us wherever we go. An image can be a photograph or a video frame. Images convey considerable amount of information about the object photographed. Objects can be 2-dimensional, such as photograph or 3-dimensional, like a statue. To be precise, an image is a 2-dimensional snapshot of the object photographed by a camera. It becomes very important to retrieve the information from this image correctly. Taking measurements on the image will not give correct results of the original object. For reliable measurements, it is wise to take measurements on the original objects.

Metrology is the scientific study of measurements. It is the act of assigning a particular value to a physical variable. Measurements provide a basis for opinion on processing information, quality control, and process control. The choice of the measuring device chosen is also important as it depends on resolution and range of desired measurement. A dependable technology nowadays is Video Metrology, performed with video measuring systems. The availability of sophisticated video cameras has thrown more interest on metrology using computer vision techniques.

A camera captures an image, and the desired portion of image is optically magnified. Later the magnified image is converted into a video signal which is studied under various electronics and software methods to determine its features. Reasons why video metrology is prominent are – (i) video contains multiple frames. Hence it provides more reliable results. (ii) It results in effortless automation in video processing. For video metrology, there are two types of cameras in use – stationary cameras and planar motion cameras. The current video-based metrology methods are extended from image-based metrology approaches [1]. Image-based metrology has limitations like the algorithms are not open to error analysis and it depends on parallel line segments. On the other hand, video-based metrology composes of multiple frames, hence these limitations are overcome.

An interesting application of metrology is in the study of human feature extraction and measurements. The features are different measurable body parts. Every Human has some dissimilar measurements, even though other measurements are identical.

The main objective of this work is to examine whether the data provided is reliable in discriminating between people. This can be answered based on performance measures such as Precision & Recall, and ROC curves. It is critical to know the performance of a system completely before validating the system. The tangible dataset used in this work is made available through CAESAR survey [2].

1.2 Problem and Motivation

In recent years, biometric recognition has become more challenging. Biometric recognition plays a vital role in various fields like surveillance and activity monitoring, human computer interaction, intelligent environments, etc. Biometric features can't be borrowed, stolen, or forgotten, and making a false biometric feature is not easily achievable. Some popular methods for biometric recognition include iris scan, fingerprint recognition, face recognition, voice recognition, etc. These methods were extensively studied and used widespread. A brief description of each biometric recognition method can be found in [3]. Some methods related to human metrology are briefed here.

Facial recognition is to identify or verify a person automatically from an image or video frame. This is done by comparing the selected facial features from that person and comparing with facial database. Every face has numerous distinguishable landmarks. These are the different peaks and valleys making up facial features. Some features are distance between the eyes, width of the nose, shape of cheekbones, etc. Each human face consists of around 80 nodal points [4]. All these nodal points are measured resulting in a numerical code known as faceprint [4], which is that person's face in the facial database. This recognition method is non-intrusive and a cheap technology. However it suffers from flaws when there is a lighting change, or person covers his face (mask), or growth of facial hair with age.

Hand Geometry recognition is one of the longest implemented biometric recognition methods. The recognition system constitutes measuring and recording the length, width, thickness, and surface area of the person's hand while placed on a scanner. These systems use a camera to capture a silhouette image of the placed hand. They are a popular choice because of their ease of use, stand-alone capabilities, and less data requirements. It can be easily integrated into other devices/systems. However it has some constraints like the hardware is very expensive, and significant in size. Also some people can't place their hand properly on the scanner.

Gait recognition is to identify a person by recognizing the way that person walks. Much biometric advancement in this field has led to the improvement of gait recognition algorithms.

Gait has many advantages while acquiring images. The images can be obtained easily from a simple camera, even in public places and without the knowledge of the subject. But it has limitations, like we do not know the degree to which a person's gait is unique. Moreover gait can be affected by factors like footwear, fatigue, terrain, and injury.

Human identification can be done through the above recognition methods. However, the available recognition methods have their respective drawbacks. A common drawback of the human recognition methods is the subject should be willing to make the scans on him/her. This may not be the case always. Sometimes we want to know biometric details of a person without their knowledge. A major factor influencing this kind of identification is the distance of subject from camera. The person under study can be located at any distance from the camera. In such scenarios, video metrology plays a key role in human recognition.

Study of distanced objects from the camera is of profound interest to many researchers. Government and Private Agencies have carried out many related projects in this field before. For example, The Information Processing Technology Office (an agency of DARPA) ran a program known as "Human Identification at a Distance" which developed technologies that are capable of identifying a person at up to 500 ft by their facial features.

Hence it is significant to recognize humans at a distance from the camera, which is commonly observed in the real world. This thesis studies performance of different human measurement features in discriminating between people.

1.3 Contribution of the Thesis

As mentioned in the previous human recognition methods, most of these methods have the constraint that the human under study should be within the proximity of the measuring instrument. If this condition isn't satisfied, the system doesn't recognize the human due to erroneous measurements.

The principal aim of this work has been to establish performance of human metrology based on distance plots, precision & recall, and false acceptance rate & genuine acceptance rate. We have used the CAESAR database for calculating the performance measurements. From the complete dataset provided, we selected some features of both male & female subjects which are commonly used, and later feature extraction is performed to achieve uncorrelated dataset.

This thesis details the performance measurement through some ROC curves. Also some characteristic plots are shown to augment the performance of this database.

1.4 Thesis Organization

Chapter 2 is a brief overview of studies carried out before in the field of human metrology. Related research dealing with CAESAR data or statistical analysis is briefed. Also biometric system is explained in brief, along with function of metrology system.

Chapter 3 discusses about CAESAR dataset in detail. This chapter gives a description of the survey and the data collected and produced, along with the samples strategy. Later the CAESAR scan postures are illustrated along with 3-D Landmark data points.

Chapter 4 deals with statistical analysis, the variation of CAESAR data among the population. Later we study about the different pattern recognition methods we made use in this work. This covers classification models, clustering, feature extraction and feature selection.

Chapter 5 is devoted to the performance measurement of CAESAR data. Here we implement the algorithm for this measurement. Later the performance measures are discussed in detail along with the plots.

Chapter 6 concludes our study with suggestions for future directions.

Chapter 2: Background

2.1 Related Work

Object and feature identification has been an active area in machine vision learning for the past two decades. Among this, few studies focused on identification and analysis of human body motions captured by video images [5]. In 1984, Akita [6] focused on coarse recognition tasks of human body parts. In their study, each sub program which are entitled as body parts (head, legs, arms, etc), corresponds to the labeled binary output image representing different parts of the body. But most of the previous related techniques were utilizing mono imagery for recovery of non-metrical information from images.

In the recent past, Researchers have generated interest in analyzing human body shapes, their reconstruction, and search for a reliable way to cluster humans. CAESAR dataset has been used extensively before, depending on the requirements. For example, Allen et al. described a template-based reconstruction strategy to establish correspondence among same structures, but which have significant deviation in shape [7]. They formulated an objective function to trade off the fit within the data, and fit high resolution template meshes for human body scans using sparse 3D markers. They demonstrated that the scans could be matched to reasonable degree. But a major drawback of their work is the pose of the template should be similar to the target position.

Also, based on user's specifications and need, it is possible to synthesize 3D human body shapes [8]. This is done using a corpus of complete 3D human body laser range scans of different people. A common template mesh is warped on each scanned image, leading to a vertex correspondence between body shapes. Later the variation of body shapes is related with tangible parameters. It also deals with generating human character models automatically.

Godil et al [9] cited static anthropometric distances to simulate computer vision identification biometric system for human identification. They designed a biometric composed of distances among rigid body parts. The advantage of this design is it is invariant to body posture.

Another effort was put in to explain a framework to similarity based retrieval and clustering from a 3D human database [10]. Four methods were developed for searching human database based on similarity of human body and head shape. These head and body descriptors fairly represent the CAESAR bodies accurately.

2.2 Biometric System

A biometric system is a pattern recognition system which recognizes a person based on a feature vector derived from specific intrinsic physical or behavioral characteristics possessed by that person. After the features are extracted, they are stored in a database. Biometric characteristics are of two types:

- **Physiological** traits which are related to shape of the body, like DNA, fingerprint, face recognition, iris recognition, odor, etc.
- **Behavioral** traits which mean the behavior of a person, like gait, voice, keystroke, signature, etc.

In terms of reliability, a biometric system based on physiological characteristics is more preferred compared to a biometric system adopted on behavioral characteristics. A biometric system operates on two modes [2]:

- *Verification* mode where system captures a biometric data of the person and compares with original stored templates of that person in the system database. This is done to verify that the individual is who he claims to be. This is generally done using a smartcard, PIN, username or ID number.
- *Identification* mode where the captured biometric data of a person is compared against templates corresponding to all users. This is done to identify an unknown individual. The identification of unknown person takes place only if the comparison is within a set threshold.

A human characteristic can be used for recognition through a biometric system in terms of some parameters. Biometric systems must be continuously evaluated and verified in order to be accepted. Biometrics is the core component in a video metrology system. The basic functionality of a biometric system is retained in human recognition through video measurement.

2.3 Metrology system

As mentioned before metrology is the science of measurement. It includes all theoretical and practical aspects of measurement. It is always good to get correct measurements, because wrong or inaccurate measurements can lead to wrong decisions. This can have serious consequences, costing money and sometimes lives. It is important to have reliable and accurate measurements approved by relevant authorities worldwide. Also the ever increasing demand for greater accuracy and increased reliability should be satisfied.

An important concept in metrology is traceability. It is the property of the result of a measurement which can be related to stated references. The level of traceability establishes the level of compatibility. Traceability is obtained by calibration, thereby establishing a relation between the measuring instrument and measurement standard. Calibration is the process where metrology is applied to measuring equipment and ensures conformity with a known standard of measurement. These standards are coordinated by National Laboratories.

Metrology as a science of measurement attempts to validate the data obtained from test equipment. Practically, it is the verification and validation of predefined standards for these purposes:

- Accuracy – it is the degree to which final product corresponds to measurement standard
- Precision – is ability of a measurement to be reproduced consistently
- Reliability – is consistency of accurate results over consecutive measurements
- Traceability – it is validation that measurement of product conforms to some standards

Accuracy is critical in determining whether a system meets desired requirements. The way system responds can be characterized by two error statistics: *False Accept Rate (FAR)* and *False Reject Rate (FRR).*

False Accept Rate is the percentage of imposters accepted, and False Reject Rate is percentage of authorized users rejected. They will be discussed in detail in Chapter 5 of the thesis.

Chapter 3: CAESAR Dataset

3.1 Introduction

This chapter gives a general description of the survey and the data collected and produced. This data is the principal source of input for this work. Based on some features extracted from these measurements, we determine the performance of human metrology. Materials for this chapter are taken from [11]. More details are provided there.

The Civilian American and European Surface Anthropometry Resource (CAESAR) project was a survey of the civilian populations of three countries representing the North Atlantic Treaty Organization (NATO) countries – The United States of America (USA), The Netherlands, and Italy. The Survey was carried out by US Air Force in alliance with other organizations. The Civilian population of these three countries was sampled in an effort to characterize the population of NATO countries as a whole. The main reason for choosing these populations is because of the diversity the three countries offer – United States has the largest and most diverse population, Netherlands has the tallest population, and Italy has one of the shortest populations.

Typically, the principal product from an anthropometric survey is a document with summary of population statistics, often included by means, standard deviations and percentiles. But for engineering applications just this information is not sufficient. In practical applications we need 3-D data which cannot be summarized from means, standard deviations and percentiles. To satisfy these requirements, CAESAR project was introduced. The outcome is the raw data, including for the first time ever complete 3-D models of all subjects. Thus the product of CAESAR survey was to characterize anthropometric variability of populations, including complete 3-D models of all the subjects.

3.2 Sampling Strategy

The populations were sampled by age, race, and gender. The reason for using age as strata was to ensure that all racial groups' body sizes and shapes are adequately represented. A stratified sampling plan was used with equal sample size in each cell according to ISO recommendations. There are total 42 sampling cells; 18 in North America, 12 in The Netherlands, and 12 in Italy. The minimum sample size for each cell was calculated based on the following formula:

$$\frac{|\bar{X} - \upsilon| * \sqrt{n_i}}{\sigma} = \zeta \qquad (1)$$

where ζ = eccentricity, σ = standard deviation, n_i = sample size, υ = true mean of the group, $\bar{X}$ = sample mean of the group, and $|\bar{X} - \upsilon|$ = desired within cell accuracy.

The total number target for sample in one country was sum of sample sizes in subgroups. Stature measurement was used to estimate the sample size of that subject. This is because stature measure gives the most conservative estimate. A reasonable within cell standard deviation estimate for stature is 70 mm. The desired within cell accuracy was set to 10 mm.

The calculation of within cell sample size now becomes:

$$\frac{|10| * \sqrt{n_i}}{70} = 1.96$$

Which results in n_i = 188. This value is set as target number of subjects per cell. It is the number to provide a sample mean value that is within 10 mm of true population mean with a confidence of 95%. The following tables show number of subjects studied in each country for different age groups and races.

	Female					Male			
Age	18-29	30-44	45-65	Sum		18-29	30-44	45-65	Sum
White	188	373	394	955		191	353	320	864
Black	61	48	38	147		39	52	25	116
Other	58	56	37	151		51	56	30	137
Sum	307	477	469	1253		281	461	375	1117
Total Sum					2370				

Table 1: Total Number of subjects in each stratum in North America

	Female					Male			
Age	18-29	30-44	45-65	Sum		18-29	30-44	45-65	Sum
Dutch	167	200	177	544		156	152	172	480
Other	41	48	58	147		29	23	32	84
Sum	208	248	235	691		185	175	204	564
Total Sum					1255				

Table 2: Target Number of subjects in each stratum in The Netherlands

	Female					Male			
Age	18-29	30-44	45-65	Sum		18-29	30-44	45-65	Sum
Italian	252	67	57	376		235	103	50	388
Other	5	4	1	10		14	7	1	22
Sum	257	71	56	386		249	110	51	410
Total Sum					796				

Table 3: Target Number of subjects in each stratum in Italy

Now we get the total number of subjects in CAESAR dataset from three countries including both genders as:

North America = 2370; The Netherlands = 1255; Italy = 796; Total = 4431

In North America, 12 locations were shortlisted for data collection. These locations were selected in proportion to distribution of population in each of 4 regions during that time.

3.3 CAESAR Scan Postures

The populations were sampled by age, race, and gender. The products from CAESAR survey consisted of raw files and documentation like demographic data of each subject, 3-D models and landmarks in different postures for each subject, any text files with notes on subject anomalies, summary reports, etc.

Each subject was scanned in three different postures. Pose A is a standing posture. Pose B is a seated posture where the subject assumes a “comfortable working posture”. Pose C is another seated posture in which the subject raises his/her arms and head to provide the greatest possible scan coverage.

From these poses, different measurements are obtained. Traditional-style dimensions are measured on right side of body for following body parts: shoulder, ankle, arm, buttock, elbow, eye, foot, hand, knee, thigh, wrist, and scapular and triceps skin folds. For all the measurements, the investigator makes sure the subject is suitably positioned to get an accurate reading.

There are other measurements taken from 3-D scans. The 3-D scans were processed to combine the information from different scan head within a scan into one object. This results in a complete model for each pose. High-resolution measurements of body surfaces were made using a new data collection technology – 3D surface anthropometry. This technology captures thousands of points in few seconds. It provides details about the surface shape as well as 3D locations of measurements relative to each other. The resulting scan is independent of the person, making it convenient to standardize.

3.4 3-D Landmark Data

Before scanning, the subject's body is marked with 72 Anthropometry landmarks using stickers for later identification. 12 stickers were 3-D stickers which are truncated square pyramids in shape. The rest were white circular paper stickers, 12 mm in diameter. In each pose (A, B, and C), the body was marked with these stickers, as shown in Figure 1, before scanning.

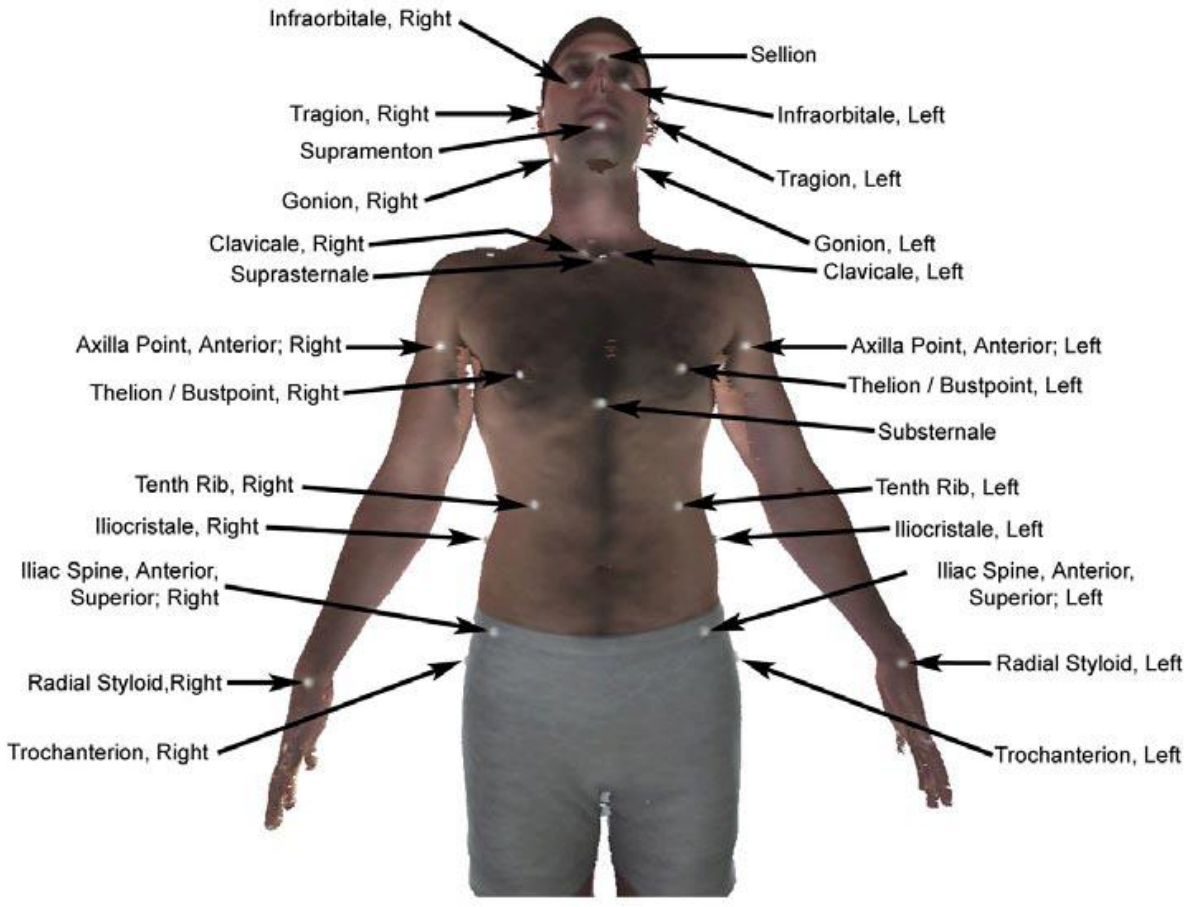


Figure 1a: Visual Index of 3-D Landmarks in Pose A, Upper Body, Front View

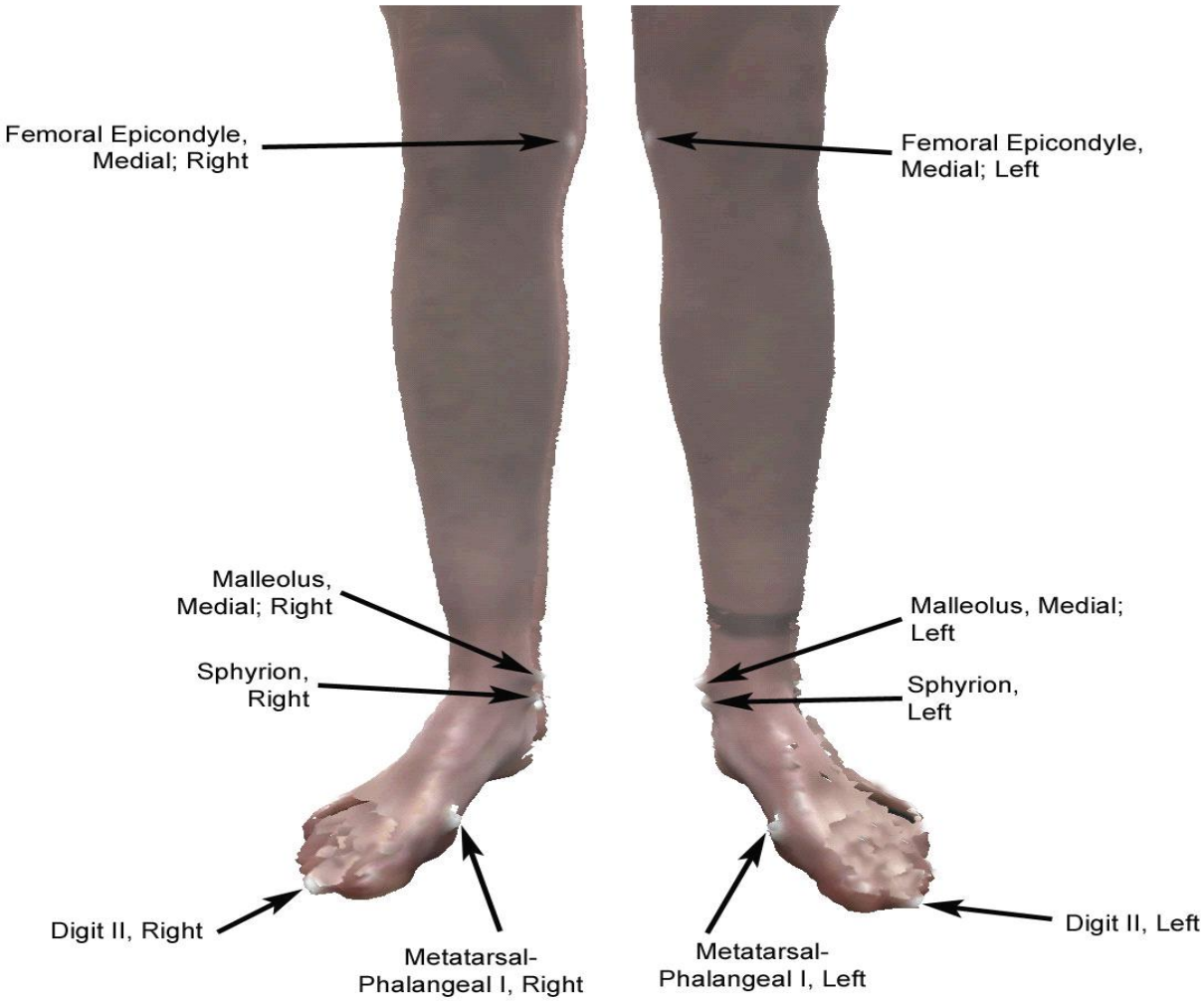


Figure 1b: Visual Index of 3-D Landmarks in Pose A, Lower Body, Front View

The Complete Landmark points for measurement in CAESAR dataset is shown in Table 4a [8]. Table 4b lists 44 measurements we have considered for this study.

1. Sellion	38. Rt. Dactylion
2. Rt. Infraorbitale	39. Rt. Ulnar Styloid
3. Lt. Infraorbitale	40. Rt. Metacarpal-Phal. V
4. Supramenton	41. Lt. Acromion
5. Rt. Tragion	42. Lt. Axilla, Ant
6. Rt. Gonion	43. Lt. Radial Styloid
7. Lt. Tragion	44. Lt. Axilla, Post
8. Lt. Gonion	45. Lt. Olecranon
9. Nuchale	46. Lt. Humeral Lateral Epicn
10. Rt. Calvicale	47. Lt. Humeral Medial Epicn
11. Suprasternale	48. Lt. Radiale
12. Lt. Calvicale	49. Lt. Metacarpal-Phal. II
13. Rt. Thelion/Bustpoint	50. Lt. Dactylion
14. Lt. Thelion/Bustpoint	51. Lt. Ulnar Styloid
15. Substernale	52. Lt. Metacarpal-Phal. V
16. Rt. 10th Rib	53. Rt. Knee Crease
17. Rt. ASIS	54. Rt. Femoral Lateral Epicn
18. Lt. 10th Rib	55. Rt. Femoral Medial Epicn
19. Lt. ASIS	56. Rt. Metatarsal-Phal. V
20. Rt. Iliocristale	57. Rt. Lateral Malleolus
21. Rt. Trochanterion	58. Rt. Medial Malleoulus
22. Lt. Iliocristale	59. Rt. Sphyrion
23. Lt. Trochanterion	60. Rt. Metatarsal-Phal. I
24. Cervicale	61. Rt. Calcaneous, Post
25. 10th Rib Midspine	62. Rt. Digit II
26. Rt. PSIS	63. Lt. Knee Crease
27. Lt. PSIS	64. Lt. Femoral Lateral Epicn
28. Waist, Preferred, Post	65. Lt. Femoral Medial Epicn
29. Rt. Acromion	66. Lt. Metatarsal-Phal. V
30. Rt. Axilla, Ant	67. Lt. Lateral Malleolus
31. Rt. Radial Styloid	68. Lt. Medial Malleolus
32. Rt. Axilla, Post	69. Lt. Sphyrion
33. Rt. Olecranon	70. Lt. Metatarsal-Phal. I
34. Rt. Humeral Lateral Epicn	71. Lt. Calcaneous, Post
35. Rt. Humeral Medial Epicn	72. Lt. Digit II
36. Rt. Radiale	73. Crotch
37. Rt. Metacarpal-Phal. II	74. Functional Butt Block

Table 4a: Landmark Points for measurement

	Mean (mm)	Std
1. Acromial Height, Sitting	586.270	37.656
2. Ankle Circumference	253.507	21.065
3. Spin-to-Shoulder	204.407	17.208
4. Spine-to-Elbow	530.339	35.958
5. Arm Length (Spine to Wrist)	817.026	57.865
6. Arm Length (Shoulder to Wrist)	612.618	46.005
7. Arm Length (Shoulder to Elbow)	325.931	24.465
8. Armscye Circumference (Scye Circ Over Acromion)	417.26	52.236
9. Bizygomatic Breadth	138.502	8.207
10. Chest Circumference	996.615	123.464
11. Buttock-Knee Length	602	39.969
12. Chest Girth at Scye (Chest Circumference at Scye)	984.383	117.991
13. Crotch Height	773.518	55.805
14. Sitting Height	893.734	48.811
*15. Stature	1704.673	102.711
16. Subscapular Skinfold	20.747	11.103
17. Thigh Circumference	607.204	69.461
18. Thigh Circumference Max Sitting	607.288	68.647
19. Thumb Tip Reach	774.221	56.254
20. Thumb Tip Reach 1	772.582	56.685
21. Thumb Tip Reach 2	775.116	56.545
22. Thumb Tip Reach 3	775.092	56.682
23. Elbow Height, Sitting	238.982	28.028
24. Eye Height, Sitting	780.235	45.173
*25. Face Length	116.530	8.537
26. Foot Length	253.07	19.915
27. Hand Circumference	197.71	17.685
28. Hand Length	192.142	14.91
*29. Head Breadth	150.763	7.26
30. Head Circumference	564.753	21.481
*31. Head Length	193.888	9.480
32. Hip Breadth, Sitting	395.957	43.966
33. Hip Circumference, Maximum	1049.87	111.802
34. Hip Circ Max Height	851.514	69.315
*35. Knee Height	534.173	39.921
36. Neck Base Circumference	437.326	40.193
*37. Shoulder Breadth	460.953	48.268
38. Triceps Skinfold	18.607	9.769
39. Total Crotch Length (Crotch Length)	676.654	71.942
40. Vertical Trunk Circumference	1645.225	129.204
41. Waist Circumference, Preferred	847.877	143.491
42. Waist Front Length	418.649	65.603
*43. Waist Height, Preferred	1025.878	59.908
44. Weight	77.014	19.660

Table 4b: 44 measurements selected for this study along with their average values and standard deviations

NOTE: * indicates those in our smaller set of 11 measurements

Based on the landmarks listed in Table 4a, there were total 99 measurements. Since many of the traditional measurements have been used for many years, it is more reliable to take some measurements the traditional way. As a result 40 measurements were taken with calipers and tape measures, and 59 measurements were point to point or point to surface, that are calculated from the scan points. In the latter measurements, 43 were calculated from 3-D landmarks from standing pose and 16 were calculated from 3-D landmarks from seated pose. The measurements are in alphabetical order within each section according to their CAESAR name. The measurements are stored in both English units (inches) and metric units (millimeters). The CAESAR name is consistent with ISO rules. The body part or point is listed first, followed by the type of measurement. This is followed by the pose, if necessary, and later followed by the side of body when both are measured (Left or Right). In the data ISO names are mentioned, along with name used in raw data file provided. This complete data is provided in both ASCII text and EXCEL spreadsheet files.

In case of any missing values in the measurement of a subject, we estimate the expected value of this measurement from this subject's remaining measurements. Here we predict any missing data based on the other measurements of a person. In the complete dataset of 2370 persons, there were 5 persons with few missing measurements.

Chapter 4: Statistical Analysis and Feature Extraction

4.1 Introduction

Data can be collected from existing sources or can be obtained from observation and experimental studies. The patterns in data can be studied for randomness in the observations and we can draw inferences about the population distribution.

The input data for this work is the CAESAR dataset. We want to look at performance of this dataset in classifying humans. First, we want to answer a few questions like what is the nature of the dataset? What is the relation of dataset in underlying population?

Probability is vital in decision making because it provides a mechanism for measuring, expressing, and analyzing uncertainties associated with the data in future events.

In this work, we approach the study of performance by considering the four-step process in learning from data [12]:

1. Defining the problem
2. Selecting the data
3. Summarizing the data
4. Analyzing the data, interpreting the analyses, and communicating the results.

All the above four steps are equally important. The CAESAR dataset has intrinsic behaviors under varying conditions. Before application of CAESAR dataset in scientific applications, it is important to have a thorough knowledge of the behavior of this data. We want to achieve our goal of performance measurement as:

- We want to summarize the whole data in a shorter form through feature extraction
- Use some basic statistical analysis methods on these features
- Design and implement an algorithm to study variations in these measurements
- Establish performance of this data through ROC curves

Understanding the statistics of a dataset thoroughly means to deal with uncertainty introduced by errors in measurement as well as by other fluctuations. Obviously performance should be of

even greater value in situations of high precision than in those in which the data are affected by large errors.

In this thesis we will be doing some analysis not only on the original data provided, but also on few variations of these original measurements with the help of varying thresholds. We also predict and forecast some features to get expected features. This is discussed in the further sections of this chapter. It is necessary to carefully plan how many features are needed from the complete dataset, and how are they extracted from the original set. Also, what are these features? These questions are answered in the next section.

4.2 Selected Measurements

The data for this work is the 99 measurements obtained from CAESAR survey. From this dataset, we initially select 11 features which are easier to extract automatically, and are more often employed. Later we consider 44 features, and performance measurement is established on both these set of features. We call these sets SET-11 and SET-44 measurements respectively.

The SET-11 measurements selected from the original dataset are listed in Table 5:

measurement name	code	landmark points	ID
body height	bh	1	(i)
eye height	eh	2&3	(ii)
chest height	ch	13&14	(iii)
elbow height	eh	33&45	(iv)
waist height	wh	28	(v)
knee height	kh	55&65	(vi)

measurement name	code	landmark points	ID
face breadth	fb	2&9	(vii)
face length	fl	1&4	(viii)
head breadth	hb	5&7	(ix)
head length	hl	1&9	(x)
shoulder breadth	sb	29&41	(xi)

Table 5: SET-11 measurements

For a more comprehensive study of the performance, we later select SET-44 measurements. These are listed in Table 4b. Some measurements present in SET-11 are also indicated with '*' in this table.

4.3 Distribution of Measurements in CAESAR dataset

For uniformity, each selected feature is normalized using two different methods. They are:

$$\text{(i) Normalized value} = \frac{(M - \min(M))}{(\max(M) - \min(M))} \quad (2a)$$

$$\text{(ii) Normalized value} = \frac{(M - \mu)}{\sigma} \quad (2b)$$

where, *M* is the feature vector, μ and σ are mean and standard deviation of feature vector M respectively, *min(M)* and m*ax(M)* are the minimum and maximum values of vector M respectively. Each feature is normalized against its own values.

Before performance analysis, it may be informative to analyze the data at hand. We want to observe distribution of the dataset and discover any related patterns. This can be achieved using scatter plots and histogram distributions.

4.3.1 Pair-wise Scatter Plots

Scatter plots are bivariate or trivariate plots of variables against each other. They aid in understanding relationships among the variables of a dataset. A downward sloping scatter plot indicates that as we increase the variable on the horizontal axis, the variable on vertical axis decreases. An analogous statement can be made for upward sloping scatters.

Figure 2 shows the pair wise scatter plots between all SET-11 measurements along with a best-fit trend line (marked in red) for the entire population of 2370 subjects. A total of 110 scatter plots were obtained; 10 for each distribution of one feature against all other features. The horizontal and vertical axes range from 0 to 1 due to normalization using Equation 2a.

The plots were used to get a basic idea of the distribution of these easily extractable features. We can know whether the data is displaying an accelerating, decelerating, or stationary trend. Also a “best-fit” trend line is applied to the data points. The proportion of data points above and below the trend line shows variation from the expected values. The initial step in statistical analysis will be to find the correlation and covariance of the provided

data. From scatter plots we can see that some features are well correlated, and some are uncorrelated. This shows randomness in measured data.

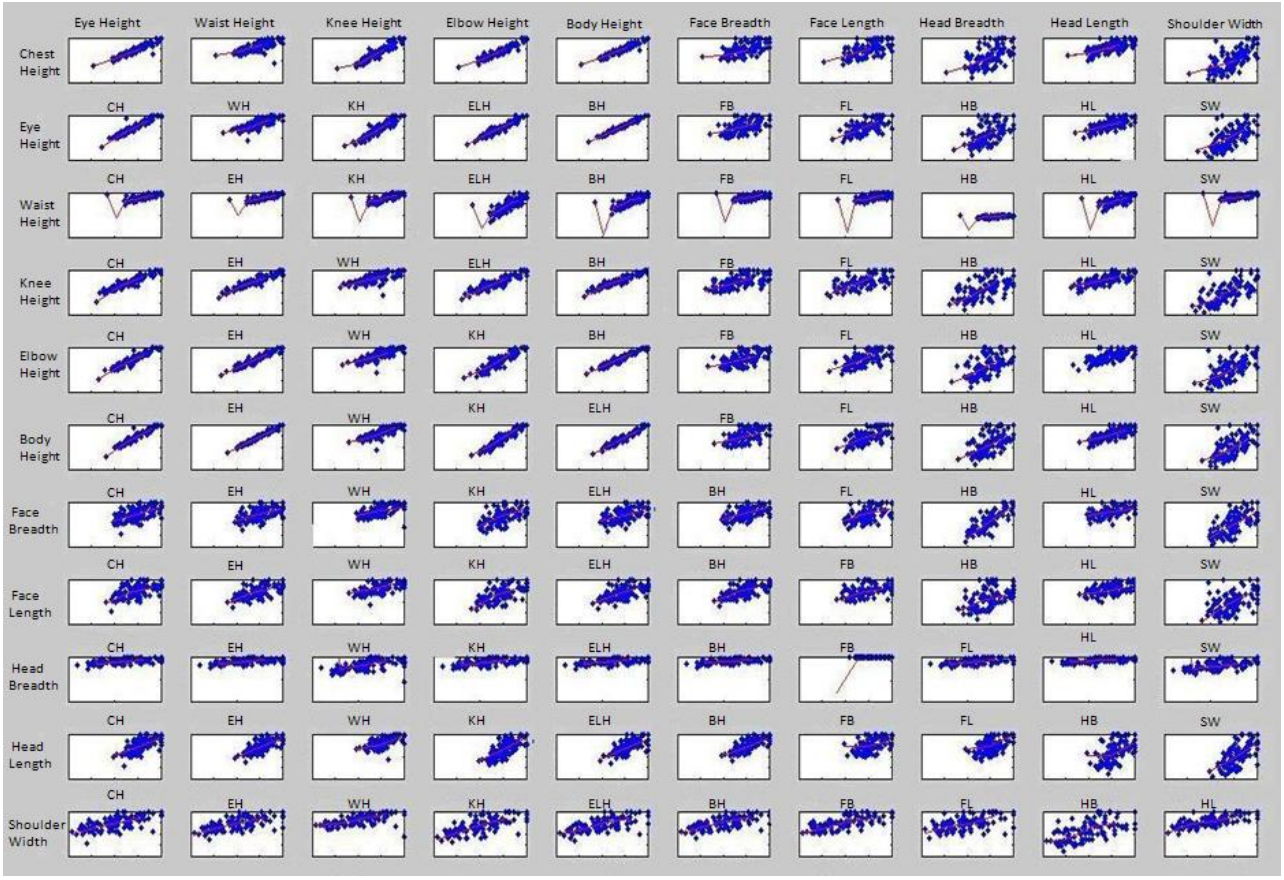


Figure 2: Pair-wise scatter plots for SET-11 measurements over entire population

4.3.2 Histogram Distribution

For each measurement, a histogram is also plotted to give more detail about the measurements variation in the population. Figure 3 shows the histogram distribution of SET-11 measurements over the entire CAESAR population of n = 2370 subjects.

From the histogram below, we get an idea of the shape of distribution of this large dataset. They are useful in describing large differences in shape and symmetry. However they can't be used in more precise judgments such as depicting individual values.

Scatter plots and Histograms provide a basic idea of distribution of the measurements in the CAESAR dataset. But to perform further analysis, we have to think of other methods. Visual

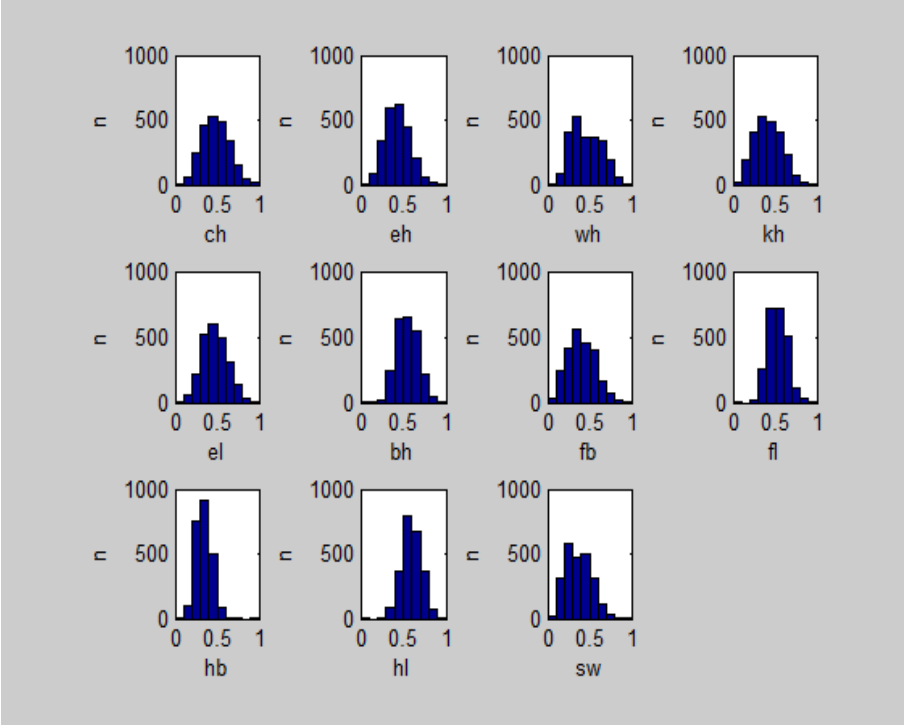


Figure 3: Histogram distribution for SET-11 measurements over entire population

inspection was the method of analysis used for these two plots, along with computation of covariance and correlation of this dataset.

4.4 Mean and Correlation values

Over the complete CAESAR data, the average values of the SET-11 features chosen initially are shown in Table 6. The average value of SET-44 features is shown in Table 4b.

Feature	Average value (mm)
Body Height (bh)	1704.673
Eye Height (eh)	1470.431
Chest Height (ch)	1230.884
Elbow Height (el)	1099.438
Waist Height (wh)	1025.878
Knee Height (kh)	534.173
Shoulder Breadth (sb)	460.953
Face Length (fl)	116.530
Face Breadth (fb)	138.43

Head Length (hl)	193.888
Head Breadth (hb)	150.763

Table 6: Average value in mm of SET-11 measurements

Correlation

Pearson product-moment correlation coefficient is the most familiar quantity to measure covariance between two variables. The correlation coefficient $\rho_{X,Y}$ between two random variables *X* and *Y* is given by

$$\rho_{X,Y} = \operatorname{corr}(X,Y) = \frac{\operatorname{cov}(X,Y)}{\sigma_X \sigma_Y} = \frac{E[(X-\mu_X)(Y-\mu_Y)]}{\sigma_X \sigma_Y} \quad (3)$$

where, μ_X & μ_Y are mean or expected values of variables *X* and *Y* respectively, σ_X & σ_Y are corresponding standard deviations of variables *X* and *Y*, E is the expected value operator, *cov* is covariance, and *corr* is denoted for correlation.

Correlation is a statistical measurement of relationship between two variables. Correlation values range from +1 to -1. A zero correlation indicates there is no relationship between the features. That means the features are uncorrelated. A correlation value of -1 represents perfect negative (decreasing) linear relationship and a correlation value of +1 represents perfect positive (increasing) linear relationship. Correlations are useful because they indicate a predictive relationship which is useful in understanding about data variation [13].

In MATLAB, correlation is an inbuilt function, denoted by *corr*(*X,Y*) where *X* and *Y* can be any of the SET-11 or SET-44 measurements. Function *corr*(*X,Y*) returns a p_1-by-p_2 matrix containing the pair wise correlation coefficient between each pair of columns in the n-by-p_1 and n-by-p_2 matrices of *X* and *Y*. The correlation matrices of SET-11 and SET-44 measurements are shown in Table 7 and Table 8 respectively.

	bh	eh	ch	el	wh	kh	sb	fl	fb	hl	hb
bh	1.00	0.985	0.965	0.961	0.778	0.947	0.634	0.608	0.475	0.623	0.219
eh	0.985	1.00	0.964	0.962	0.762	0.943	0.617	0.593	0.453	0.598	0.1943
ch	0.965	0.964	1.00	0.95	0.737	0.942	0.59	0.562	0.433	0.58	0.165
el	0.961	0.962	0.95	1.00	0.757	0.918	0.631	0.576	0.476	0.597	0.221
wh	0.778	0.762	0.737	0.757	1.00	0.722	0.607	0.539	0.511	0.591	0.309

kh	0.947	0.943	0.942	0.918	0.722	1.00	0.616	0.594	0.461	0.599	0.196
sb	0.634	0.617	0.59	0.631	0.607	0.616	1.00	0.559	0.673	0.595	0.360
fl	0.608	0.593	0.562	0.576	0.539	0.594	0.559	1.00	0.520	0.589	0.388
fb	0.475	0.453	0.433	0.476	0.511	0.461	0.673	0.52	1.00	0.47	0.615
hl	0.624	0.598	0.58	0.597	0.591	0.599	0.595	0.589	0.47	1.00	0.409
hb	0.219	0.194	0.165	0.221	0.309	0.196	0.36	0.388	0.615	0.409	1.00

Table 7: Correlation matrix of SET-11 measurements

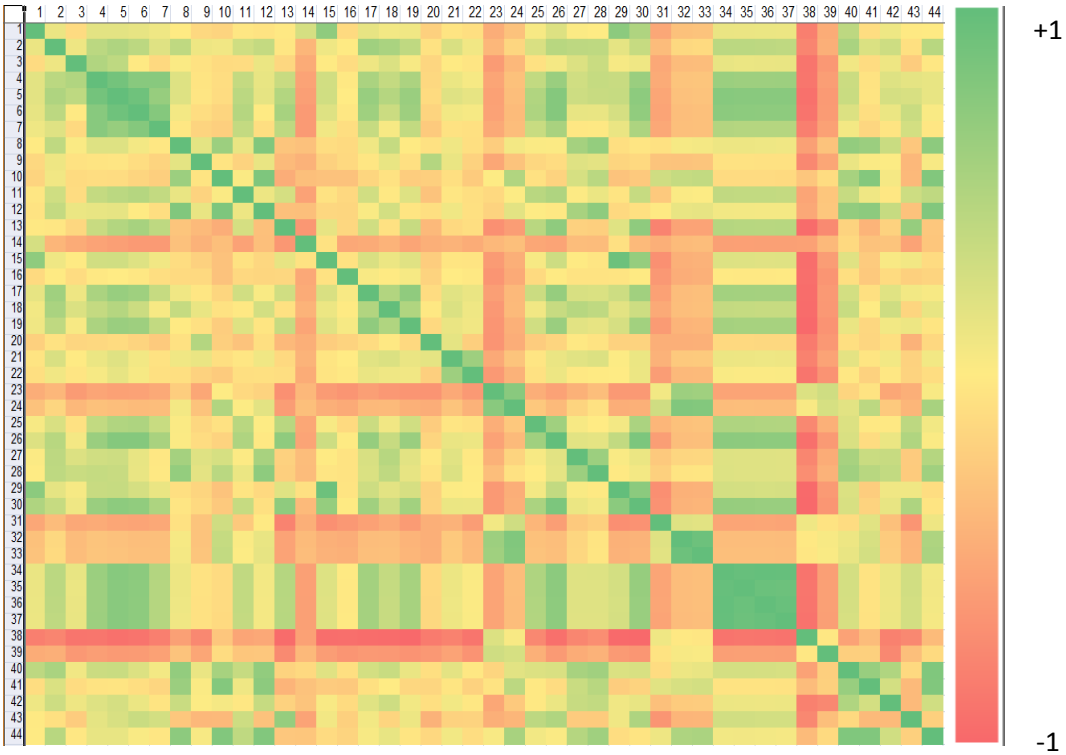


Table 8: Correlation matrix of SET-44 measurements

From the pair-wise scatter plots in Figure 2, we get an idea of relations in the chosen measurements. We observe that some measurements have linear dependence, while some are randomly distributed.

For example if we consider body height and elbow height or eye height, the plots show a linear dependence between them. This is because for tall persons, it is expected to have more eye height or elbow height. These features are highly correlated. Moreover, if we look at the head/face variation with body height, they are randomly distributed. This is due to the reality that a tall person needn't have all his measurements large, or vice versa. The head/face

measurements are independent of the body height measurement, and there is less correlation between such features.

4.5 Feature Extraction

In most practical applications, the original input variables are preprocessed to transform them into some new space of variables. This reduces variability within each class and makes pattern classification much easier. This pre-processing stage is called feature extraction. Pre-processing is also used sometimes to speed up computation e.g., using only selected extracted features. Care must be taken during preprocessing because it is possible that important information is discarded. This results in decreasing the accuracy of the system. Even our work is based on 99 features obtained from CAESAR database. We need to do feature extraction as many features can be redundant. As mentioned before in Section 4.3, we normalize the complete dataset using two normalization methods as in Equation (2). We first select 11 measurements from the original dataset. From these features, we extract other features as follows:

(i) Difference from global mean for each feature – feature vector with 11 variables

$$d_i^{global} = f_i - \mu_i, \quad i \in [1, 11]$$

where f_i is one of the feature in SET-11 and μ_i is global mean of that particular feature

(ii) The measurements are grouped into sub-groups, say 10, based on their values. We compute the differences of sub-groups from corresponding local means – feature vector with 11 variables

$$d_i^{local} = f_i - \mu_k^{local}, \quad i \in [1, 11], \ k \in [1, 10]$$

where f_i is one of the feature in SET-11 and μ_k^{local} is the local mean of k^{th} group.

(iii) Given one reference feature, we compute the expected value of other features. The expected value is obtained from pair wise relations using best-fit trendline. Now we calculate the difference between actual feature and expected/predicted value of that particular feature – feature vector with 55 variables

$$d_{ij} = f_i - f_i^{exp}(j), \ , \quad i \in [1, 11], \ j \in [1, 2370]$$

where f_i is one of the feature in SET-11 and $f_i^{exp}(j)$ is expected value of i^{th} feature and person j. We thus have a feature vector consisting of SET-11 original features and 77 (11+11+55) extracted features, making a total of 88. On this set we carry out performance measurements.

The principal idea behind choosing the other features in such a manner is to exploit different kinds of possible variations in the selected features.

4.6 Dimensionality Reduction

Dimensionality reduction is the process of reducing a large number of features into reduced number of features using either feature extraction or feature selection. It refers to pre-processing on original dataset. Performing statistical analysis on high dimensional data faces many mathematical challenges. A major problem with high dimensional datasets is not all the measured features are important in understanding the underlying phenomena of interest. It also becomes expensive to construct predictive models with high accuracy for large datasets.

Larger datasets also face setbacks while fitting a curve along the data points. Practical applications of pattern recognition deal with high dimensionality spaces. The severe difficulty that can arise in spaces of many dimensions is sometimes called as curse of dimensionality. It becomes easy to project the data from higher dimension to lower dimension for visualization and study purpose. A major linear procedure for dimensionality reduction is the principal component analysis.

Principal Component Analysis - The data transformation method used in this work is Principal Component Analysis (PCA). It is a classical statistical method to transform any possibility of correlated data into a smaller number of uncorrelated data. It performs a linear mapping of data to lower dimensions such that the variance of data in low dimensions is maximized. The correlation matrix of original data is constructed and the eigenvectors of this matrix are computed. The eigenvectors which correspond to the largest eigen values which are the principal components are used to reconstruct a large fraction of variance of original data. PCA is a non-parametric method and its applicability is limited to the assumption that the observed dataset is linear combination of certain basis. In this work we have considered principal components which account to a variance for 99% of the variance.

Chapter 5: Performance Measurement

Performance is the criteria based on which we know whether a system is reaching its desired results or not. We want to see how good the provided metrology-based features are in distinguishing humans. First, let's discuss the algorithm used in this work.

5.1 Synthetic Variations of Metrology

Here we discuss the variations we injected for the performance measurement in order to obtain evaluation metrics similar to other biometric systems. The dataset we have is measurement of different body parts for each person. These are attributes of genuine persons. In the real world, there will be imposters trying to match identities with original persons. We want to capture this and thus introduced variations in each measurement of a person.

Consider a Person *A* with measurements:

$$M^{\mathrm{A}} = \left\{m_1^{\mathrm{A}}, m_2^{\mathrm{A}}, \dots\dots\dots m_f^{\mathrm{A}}\right\}$$

where f is number of features considered, $f \in [1, 88]$

The above vector is measurement (can be original or extracted features) of Person A with zero variations given by P_0^{A}. Let the measurements be varied by a small value δ. Then the variation of Person A will be given as

$$P_1^{\mathrm{A}} = \left\{m_1^{\mathrm{A}} \pm \delta_1^{\mathrm{A}}, m_2^{\mathrm{A}} \pm \delta_2^{\mathrm{A}}, \dots\dots\dots m_f^{\mathrm{A}} \pm \delta_f^{\mathrm{A}}\right\}, \quad 0 \le \left(\frac{\delta_i^{\mathrm{A}}}{m_i^{\mathrm{A}}}\right) \le \tau, \ \forall\, i \in [1, f]$$

τ is a threshold which controls the variations. In this work we used the range $0.05 \le \tau \le 0.20$, that is 5% to 20% variation of measurements. The actual value of the added error in this range is determined using a random number generator.

For the complete CAESAR population of n persons, we consider "N" variations of each person. Hence we have one original measurement, along with N variations of that original measurement for each person

i.e., for Person A, $P_0^{\mathrm{A}}, P_1^{\mathrm{A}}, \dots\dots\dots P_N^{\mathrm{A}}$

for Person B, $P_0^{\mathrm{B}}, P_1^{\mathrm{B}}, \dots\dots\dots P_N^{\mathrm{B}}$

. . . . .

. . . . .

. . . . .

for Person n, $P_0^{\mathrm{n}}, P_1^{\mathrm{n}}, \ldots\ldots\ldots P_N^{\mathrm{n}}$

P_0^{i} represents original measurements of the i^{th} person.

τ and *N* are the critical parameters in this work. We need to be careful in choosing the values for these parameters as they have a serious impact on the performance of this dataset. For a fixed value of τ, after some values of *N*, the variability of distance remains same over the complete database. That is, even if we increase the number of variations of each original person, there is a saturation point of N after which the variability of average distance between each person and all their respective variations in the entire dataset remains constant.

For a training set of 100 people, the result for variability of distance with N is shown in Figure 4 for different thresholds of 0.05 and 0.1.

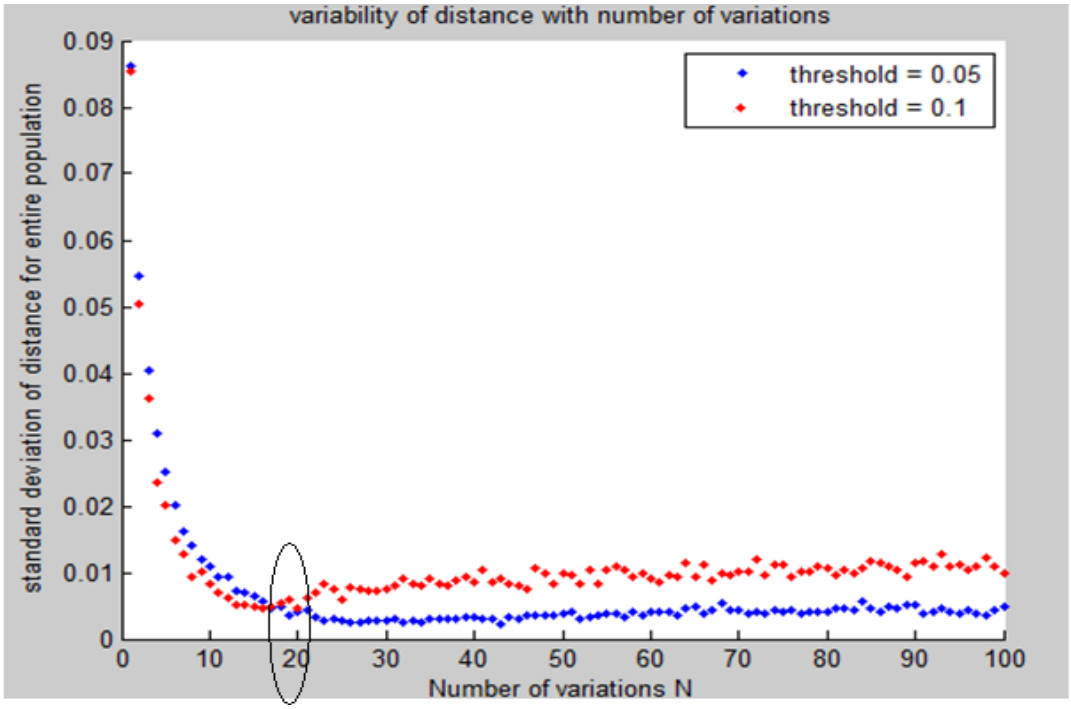


Figure 4: Variability of distance over entire population with number of variations N

As can be observed from the above figure, if the threshold increases, the standard deviation for particular value of *N* also increases and for both the thresholds, the deviation remains constant after *N* = 20.

5.2 Performance Measurement Procedure

For performance measurement of the dataset, we follow these steps:

1. Generate synthetic variations of measurements of each person.
2. We compare variations of a person with variations of other people, including original people. We store the comparisons as a distance matrix.
3. Using the distance matrix, we compute precision & recall; genuine acceptance rate & false acceptance rate.
4. We compute Eigen Person, Median Person, and Average Person for both Male & Female groups.
5. For the above mentioned Persons, we plot classification rate as a function of distance threshold.
6. We also see distribution of error rates and observe genuine persons and imposters through frequency-distance plots.

5.3 Eigen Person, Median Person and Average Person

We divide the dataset based on gender and compute eigen Person, median Person, and average Person. These are defined as follows:

Eigen Person – Eigen person is obtained from set of eigenvectors derived from covariance matrix. Eigen person can be generated by performing PCA on the large CAESAR dataset containing different body measurements. The Eigen person is person with measurements corresponding to eigen value of each measurement.

Median Person – Median person is a single person whose measurements are median values of measurements of all persons in dataset.

Average Person – Average person represents measurements corresponding to average values of all the measurements over the complete dataset.

Given two people, say A and B, we can then evaluate how similar they are by considering their respective distances from say the eigen person. Thus we compute 4 distances as:

$$\begin{aligned} d_e(A,B) &= d\{d(A,EP_K), d(B,EP_K)\}, \\ d_m(A,B) &= d\{d(A,MP_K), d(B,MP_K)\}, \\ d_a(A,B) &= d\{d(A,AP_K), d(B,AP_K)\} \end{aligned} \tag{4}$$

$$and\ d(A,B) = d(P^A, P^B)$$

where $d()$ represents euclidean distance between the specified persons; EP_K, MP_K, and AP_K are eigen person, median person, and average person respectively, and *K* = *M* or *F* to represent Male or Female gender, respectively. $d(A,B)$ is the euclidean distance between P^A, P^B.

Here all the 4 metric properties are satisfied for each of the above distances. They are: (i) $d(A,B) \geq 0$ (non-negativity), (ii) $d(A,B) = 0$ if and only if $A = B$ (identity of indiscernibles), (iii) $d(A,B) = d(B,A)$ (symmetry), and (iv) $d(A,C) \leq d(A,B) + d(B,C)$ (triangle inequality). Using the eigen Person, median Person, & average Person, and the features, we can then measure the performance of metrology-based features in human classification and identification.

5.4 Performance Measures

Here we define the performance measures of human metrology which facilitate person identification:

5.4.1 Distance Plots

For the SET-11 measurements and their extracted features as discussed in Section 4.5, we compute euclidean distance over the complete CAESAR dataset. For a given person, the value in the distance plot shows the average distance between this individual and all other people. Since each individual is supposed to be distinct for every other person, if metrology features are effective in people discrimination, then each person should have a high average distance from every other person. Thus, higher values in the plot should indicate better performance. Similarly we get the distance plot for SET-44 measurements listed in Table 4b over the complete dataset.

We also compute distance plots for each of the following cases to exploit different possible occurrences:

Case 1: Original person against only their variaitons

Case 2: Original person against remaining original people and their variations

Case 3: Original person against their variaitons and remaining original people and their variaitons

5.4.2 Classification Rate

For the 88 features representing 11 original and 77 extracted features, we compute distances of eigen person, median person and average person based on Equation (3). Their classification rates are computed for eigen, median and average persons, the combined person and considering euclidean distances at different thresholds and plotted. Similarly, we repeat the process for SET-44 measurements.

We define the classification condition between two persons A and B as: $If\ d(A,B) > \tau$, (τ is some threshold), A and B are not similar. We call this correct classification at threshold τ.

Thus the classification rate becomes

$$\Gamma = \frac{\#\ pairs\ with\ d(A,B) > \tau}{Total\ possible\ pairs} = \frac{\#\ pairs\ with\ d(A,B) > \tau}{\left(\frac{n*(n-1)}{2}\right)} \quad (5)$$

where n is number of persons considered.

We later compute the combined distance of eigen Person, median Person, and average Person and get the combined classification rate of all these persons using the following equation:

$$d(A,B) = \left(w_e * d_e(A,B) + w_m * d_m(A,B) + w_a * d_a(A,B)\right) \quad (6)$$

where w_e, w_m, and w_a are weights such that $w_e + w_m + w_a = 1$; and $d_e(A,B)$, $d_m(A,B)$, and $d_a(A,B)$ are obtained from Equation (3). Equation (6) is also used to perform classification.

We initially chose some random weights such that their sum is 1. Based on few trial and error plots for the classification rate, we observed that the eigen person's classification rate is better compared to median person or average person, and median and average person's classification rates are somewhat similar. This directed to choose the weights w_e = 0.4, w_m = 0.3, and w_a = 0.3.

We also compute classification rate by performing optimal linear discriminant analysis (LDA) on the training set and complete set and remove any redundant features. We used the Fisher's linear discriminant analysis on the data. In this analysis, we maximize the between-class variance, and minimize the within-class variance. We also choose the best features from original SET-11 and SET-44 measurements using PCA, and calculate the classification rate for these PCA features. LDA has a potential advantage over PCA because PCA only finds direction of largest variance and it is an unsupervised technique. PCA doesn't include label information of the data, and these drawbacks are overcome in LDA.

We also compute the classification rate for different age groups and each gender (male or female), as shown in the results.

5.4.3 Precision vs. Recall

Precision and recall are widely used measures for evaluating the accuracy of a pattern recognition algorithm. Precision is the degree of mutual agreement or repeatability among a series of individual measurements, values or results. It indicates the closeness with which measurements agree with one another. A precise estimate has small bias and variance.

Precision vs. recall graph provides an immediate, visual sense of the selected category's performance. When the plotted line is in the upper-right portion of the graph, the selected system is performing well. When the plotted line is in the lower-left portion of the graph, this indicates the system's performance is poor. We define precision and recall as follows:

$$\mathrm{P}_r = \left(\frac{\#\text{ correct retrieval}}{\text{Total retrieved}}\right) = \frac{|C|}{|T|}\,, \tag{7a}$$

$$\text{and } \mathrm{R}_c = \left(\frac{\#\text{ correct retrieval}}{\text{Total correct}}\right) = \frac{|C|}{N+1} \tag{7b}$$

Based on Equation (7), we compute precision and recall for 88 features and 44 measurements and plot them against each other. For the 44 features, we introduce N variations and compare originals against variations of other persons too, by setting a distance threshold of 0.1.

5.4.4 Genuine Acceptance Rate (GAR) vs. False Acceptance Rate (FAR)

False acceptance rate (FAR) or false match rate (FMR) is the probability that the system incorrectly matches a new data to a non-matching data present in its database. It gives a measure of invalid inputs incorrectly accepted.

False rejection rate (FRR) or false non-match rate (FNMR) is the probability that the system fails in detecting a match between a new data and its matching data present in the database. It gives a measure of valid inputs incorrectly rejected.

Genuine acceptance rate (GAR) is the overall accuracy measurement of a biometric system. We have GAR = 1 – FRR. GAR and FAR can be calculated at a particular rank or threshold as follows:

$$GAR = \left(\frac{\#\text{ correct retrieval}}{\text{Total retrieved}}\right) = \frac{|C|}{|T|}\,, \tag{8a}$$

$$FAR = \left(\frac{\#\text{ falsely accepted items}}{\text{Total retrieval attempts}}\right) = \frac{|F|}{|R|} = \frac{|T| - |C|}{(N+1)*(n) - 1} \quad (8b)$$

Figure 5 shows a general distribution of these error rates. The Receiver Operator Characteristic (ROC) curve is defined as the plot of

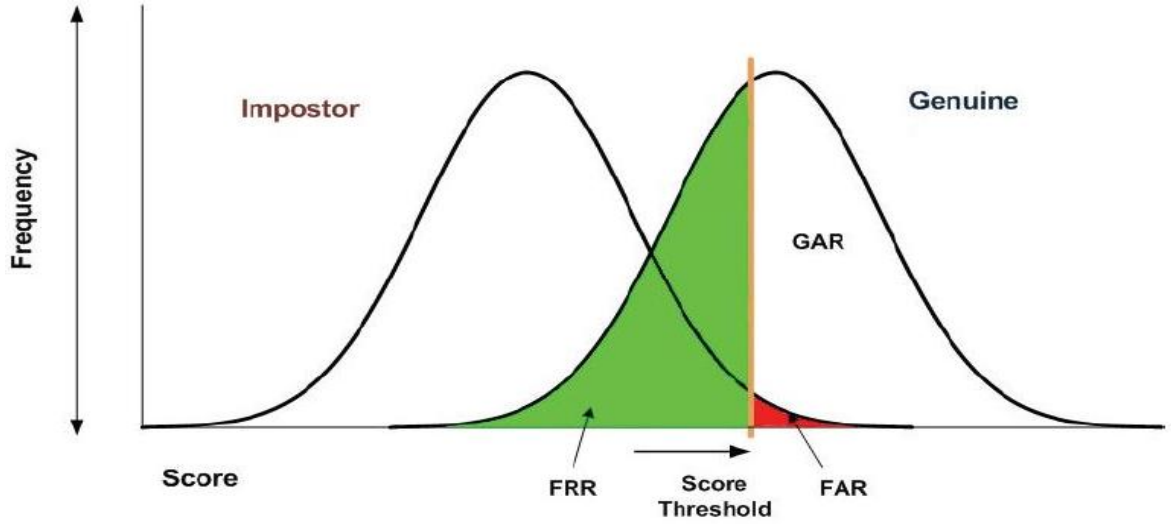


Figure 5: Relation between different error rates and their dependence on threshold

GAR against FAR. A plot between these two rates is shown in the results for both 88 features and 44 measurements.

We plot the frequency of occurrence of genuine persons and imposters. For this we consider 50 variations of each person and compare these variations with all original persons and their variations too.

For biometrics, GAR-FAR is more relevant because it helps in recognizing genuine persons and imposters. This is the key reason in designing a biometric system. Precision-recall is used in pattern recognition and machine learning to verify the correctness of an algorithm in retrieving data. The principal reason to introduce precision-recall in this work is to verify how correctly the algorithm is able to retrieve an original person's variations in a pool of many possible cases. The computational cost however depends on size of the data because more memory is needed to store large values and later test those values for correct retrieval.

5.4.5 Clustering

Clustering is an example of unsupervised learning method. Clustering deals with identifying groups, or clusters of data points in a multidimensional space. Data points in the same cluster are similar in some meaningful sense. Clusters consist of data points whose inter-

point distances are small compared to the distances to points outside the cluster. In this work we use *K*-means clustering and agglomerative clustering. We did clustering on 88 and 44 measurements as shown in the results.

We have used kmeans function in MATLAB to cluster the data. The output is a vector of indices indicating to which of the *K* clusters a point has been achieved. The method creates a single level of clusters. *K*-means computes cluster centroids differently for each measure. It uses an iterative algorithm which minimizes the sum of distances from each point to its cluster centroid for all the clusters. The algorithm moves points between clusters until the sum cannot be decreased further. This results in a set of clusters which are as compact and well separated as possible. A notable feature of *K*-means algorithm is at each iteration, every data point is assigned to one and only one cluster.

Agglomerative is a bottom-top method where each cluster is divided into sub-clusters. An advantage of this hierarchical method is we can generate smaller clusters which may be useful for further studies. But we have to know the number of clusters before hand.

5.5 Results

In this section, we show the results related to the performance measurement of human metrology. We used a training set of 100 persons (50 male and 50 female). In some cases, we also used the complete dataset of 2370 people. The plots are divided into two groups based on (i) training set and (ii) complete set of 2370 persons.

5.5.1 Distance Plots

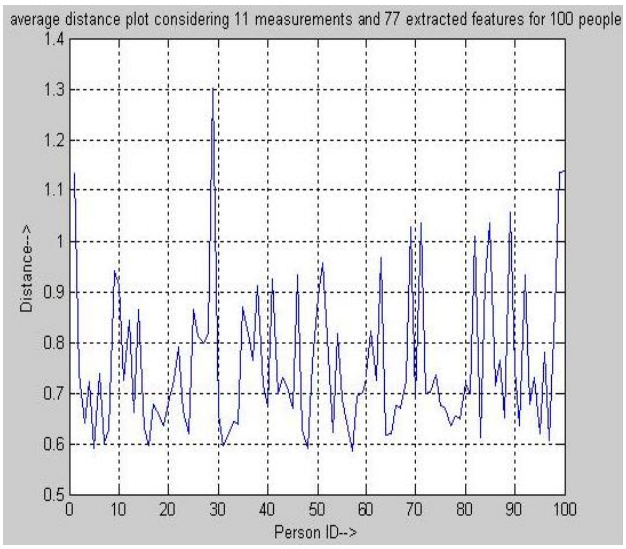
Using SET-11 and SET-44 measurements, we show the distances for originals and each of the following variaitons:

Case 1: Original person against only their variaitons

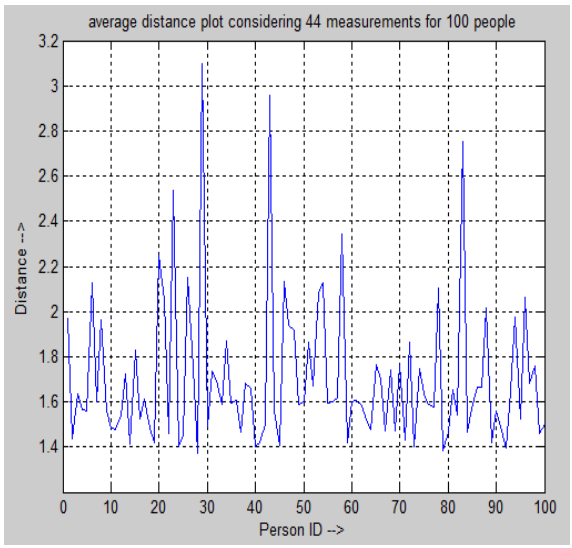
Case 2: Original person against remaining original people and their variations

Case 3: Original person against their variaitons and remaining original people and their variaitons

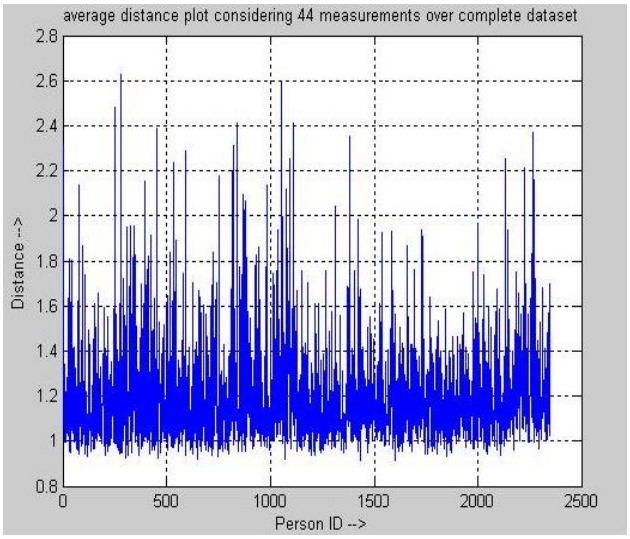
We calculate the person-to-person distance for different sets of features. The results are shown in following figures. Figure 6 shows the plots for originals.



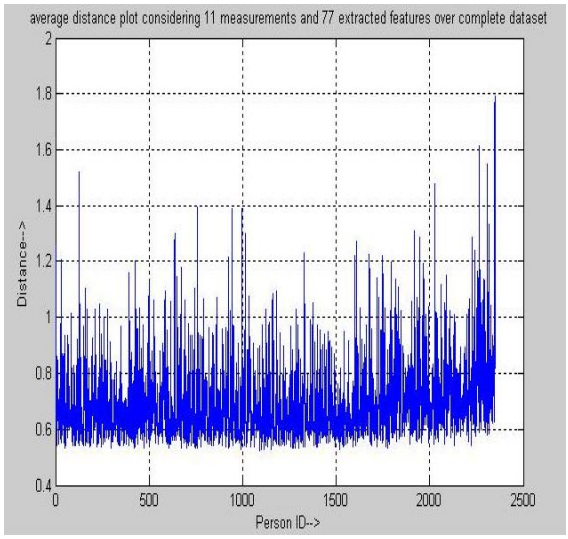
(a) 88 measurements and 100 people



(b) 44 measurements and 100 people



(c) 88 measurements over complete set



(d) 44 measurements over complete set

Figure 6: Average person-to-person distance among the population considering originals

Figure 6 shows the average person-to-person distance of original persons within the population for both training and complete set. The key observation is that the minimum distance is not zero. That is, at a certain threshold, it may be possible to separate different individuals to some extent, based on metrology features. We can also observe that the average person-to-person distance increases if the number of features considered increases. We get an idea of minimum, maximum and average distances, for each set of measurement selected, and also the number of persons. Training set contains 100 persons and complete set contains 2370 persons. The values are listed in Table 9 for both SET-11 and SET-44 measurements.

88 measurements	Max. Distance	Min. Distance	Avg. Distance	**44 measurements**	Max. Distance	Min. Distance	Avg. Distance
Training Set	1.3020	0.5853	0.7574	Training Set	3.0959	1.3718	1.7028
Complete Set	1.7979	0.5251	0.6945	Complete Set	2.6273	0.9116	1.1972

Table 9: Maximum, Minimum, and Average distance values of 44 and 88 measurements considering originals

The above table gives an idea of the max, min and avg. values taking just the originals into consideration. Comparing the originals with variations of originals is more important because in practice, there will be some imposters trying to act as an original. The distance plot for comparison of originals with variation of only the originals is shown in Figure 7.

Case 1: Original person against only their variaitons.

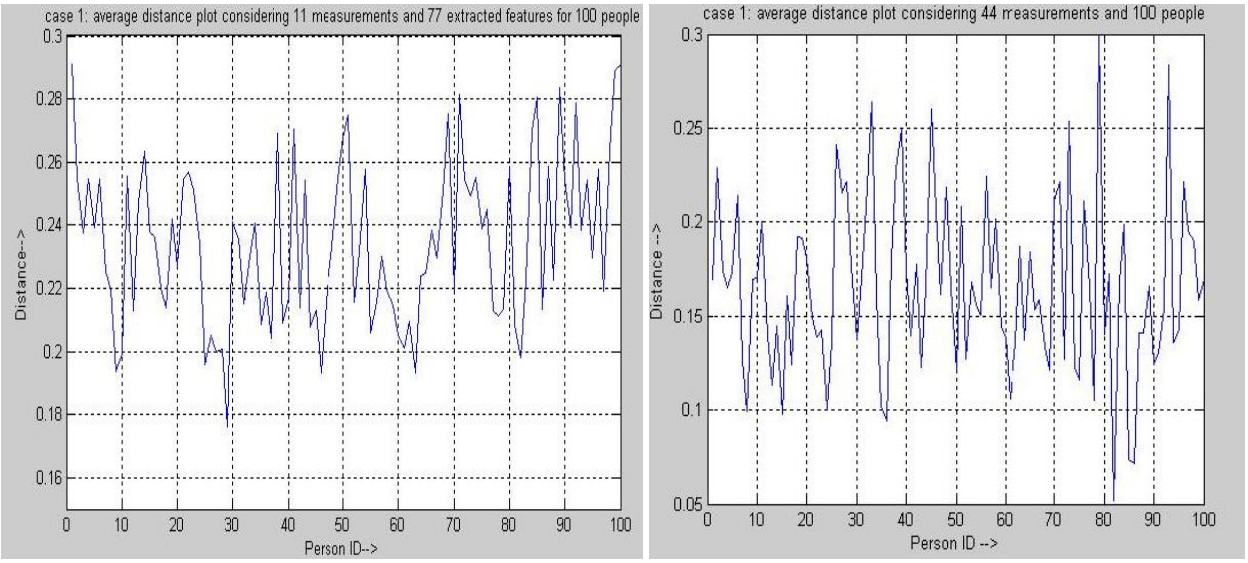


Figure 7: Average person-to-person distance for training set for Case 1

The maximum, minimum, and average value of average distances for Case 1 is shown in Table 10.

88 measurements	Max. Distance	Min. Distance	Avg. Distance	**44 measurements**	Max. Distance	Min. Distance	Avg. Distance
Training Set	0.2910	0.1761	0.2344	Training Set	0.2976	0.0519	0.1648

Table 10: Maximum, Minimum, and Average distance values of 44 and 88 measurements for Case 1

From the table, we observe that the values are very small. This is expected because the persons will be just variations of originals by τ = 0.1. We can say from the table that at a

threshold of 0.1761, all the persons (originals and their respective variations) in this case can be distinguished.

We later consider Case 2, where originals are compared with remaining originals and their variations. The distance plots are shown in Figure 8.

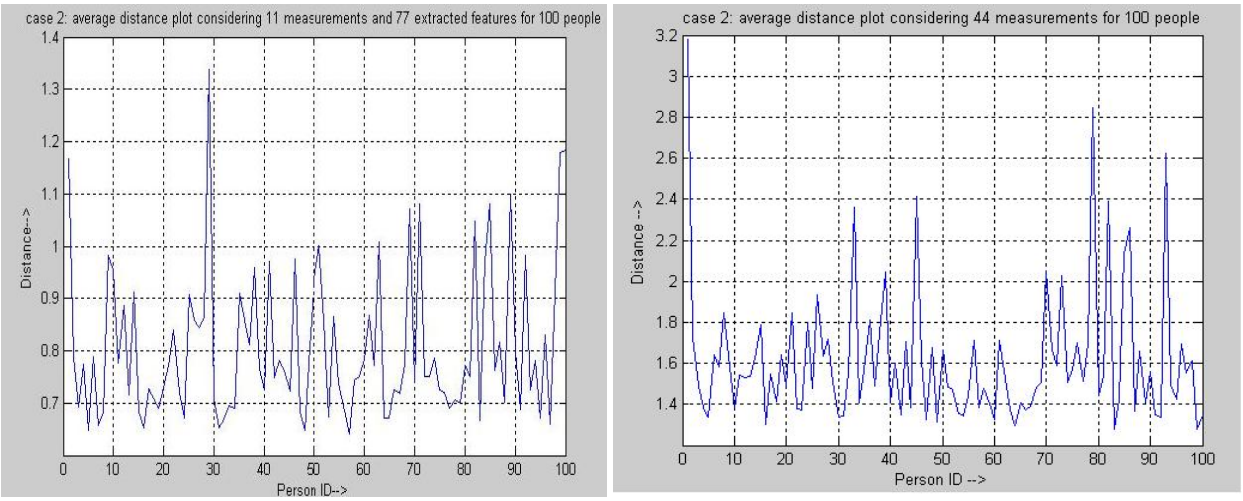


Figure 8: Average person-to-person distance for training set for Case 2

The table values are shown in Table 11.

88 measurements	Max. Distance	Min. Distance	Avg. Distance	**44 measurements**	Max. Distance	Min. Distance	Avg. Distance
Training Set	1.3369	0.6415	0.8066	Training Set	3.1817	1.2805	1.6119

Table 11: Maximum, Minimum, and Average distance values of 44 and 88 measurements for Case 2

We observe that the values increased in this case compared to Case 2. This is reasonable because here the originals are compared to originals as well as their variations.

We later consider Case 3 where originals are compared with their variations, remaining original people and their variations too. The result is shown in Figure 9.

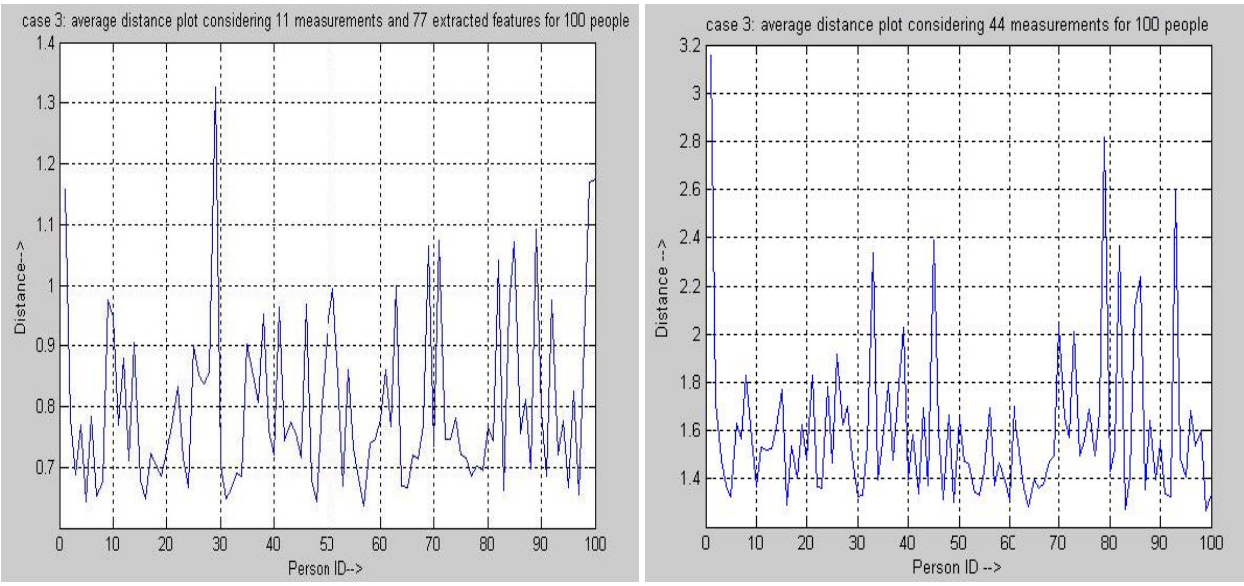


Figure 9: Average person-to-person distance for training set for Case 3

Table 12 shows the maximum, minimum, and average values of this case.

88 measurements	Max. Distance	Min. Distance	Avg. Distance	**44 measurements**	Max. Distance	Min. Distance	Avg. Distance
Training Set	1.3260	0.6373	0.8008	Training Set	3.1538	1.2694	1.5976

Table 12: Maximum, Minimum, and Average distance values of 44 and 88 measurements for Case 3

From the above table, we observe there is very small difference from values compared to Table 11. This is because of the variations of originals introduced in Case 3.

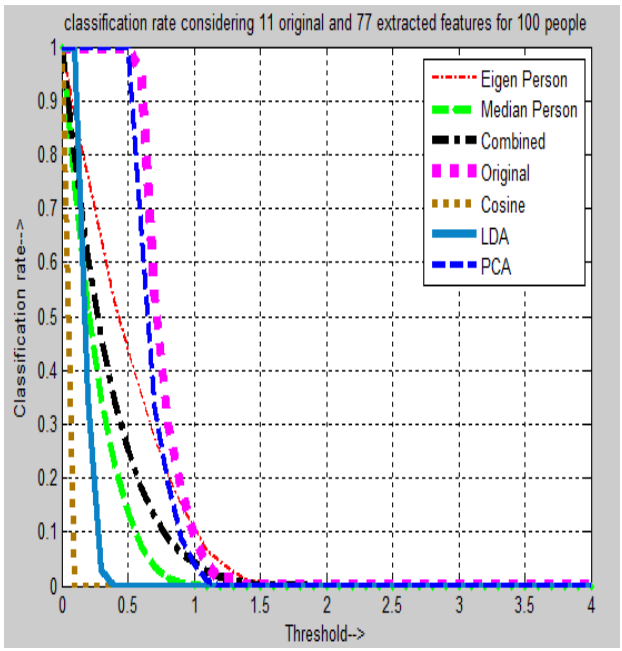
5.5.2 Classification Rate

The observations on the minimal distances above imply that we could perform some classification study by varying the thresholds. Here, each person is considered a class in itself. A classification error occurs if below a given threshold more than one person is found to be similar to a given subject. The number of such errors point to the error rate. Based on Equations (4), (5) and (6), we compute the classification rates considering the same training set, and later on the complete set. We calculate it for each case 1, 2 and 3 as like the distance plots. The results are shown below.

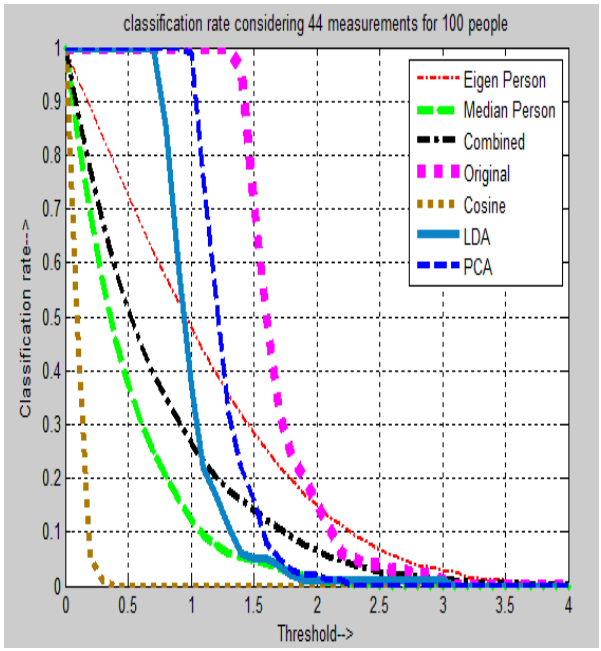
In this figure, the classification rate is shown based on euclidean distances, eigen distances, average distances, median distances, and cosine distances and after performing LDA and PCA, as discussed in Section 5.4.2. We also make use of Equation (6) to get the combined distance.

The results are shown in Figure 10(b) & 10(d) for the SET-44 measurements. As can be seen from the plots, classification rate is higher for PCA and LDA until some distance threshold and falls steeply after that.

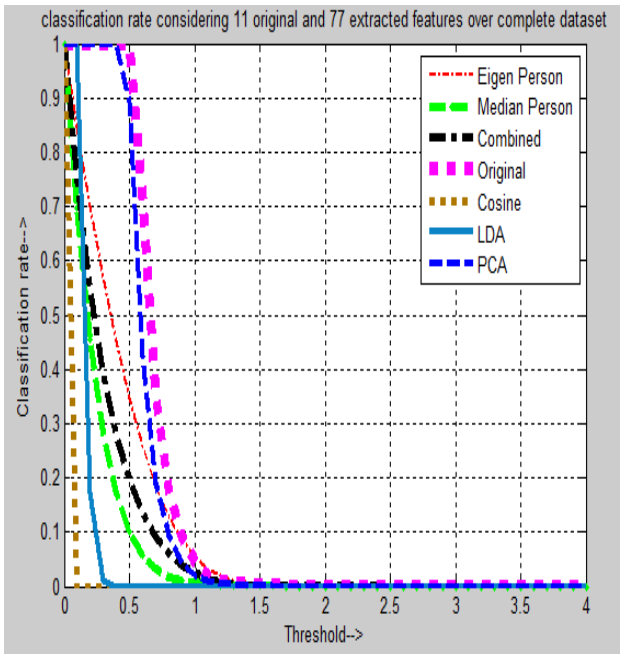
From the results we observe that euclidean distance results in the highest classification rate, and cosine distance in the lowest classification rate (Figure 10(b) & 10(d)). Because of this we exclude cosine distances for the remaining classification rate plots.



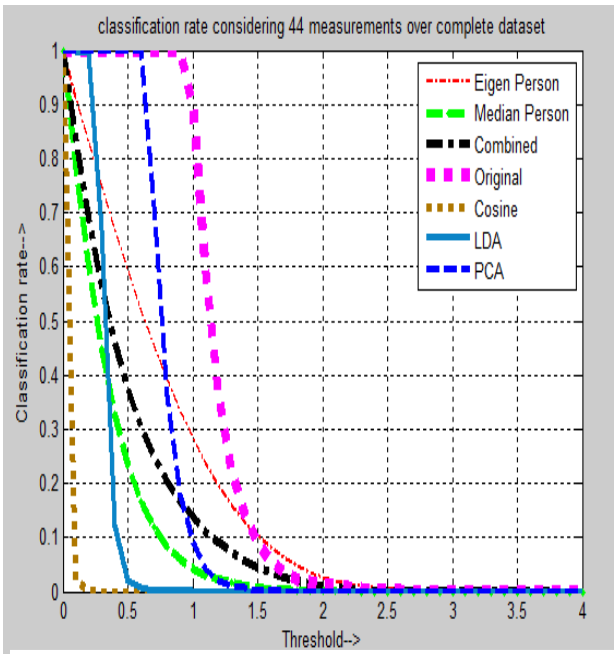
(a) 88 measurements and 100 people



(b) 44 measurements and 100 people



(c) 88 measurements over complete set

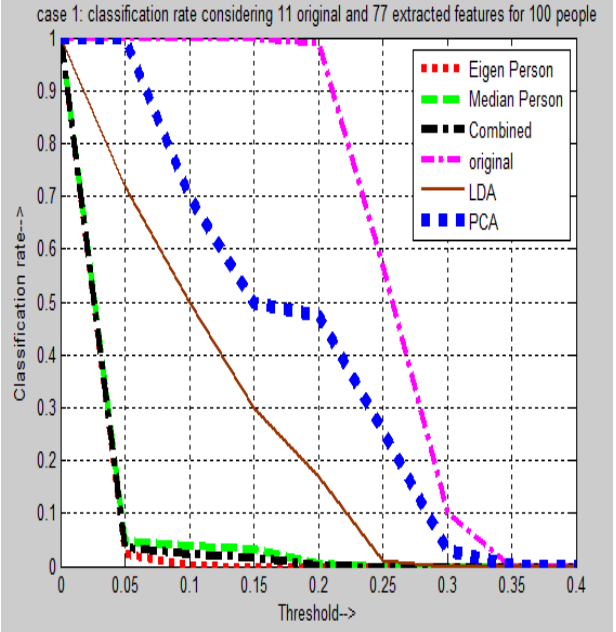


(d) 44 measurements over complete set

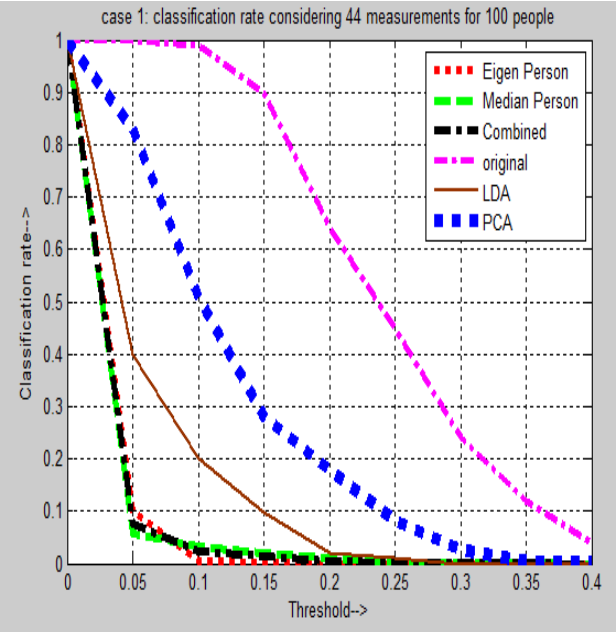
Figure 10: Classification rate as a function of threshold for original persons

Figure 11 shows the classification rates for each of the cases 1, 2 and 3.

Case 1: original person against only their variations

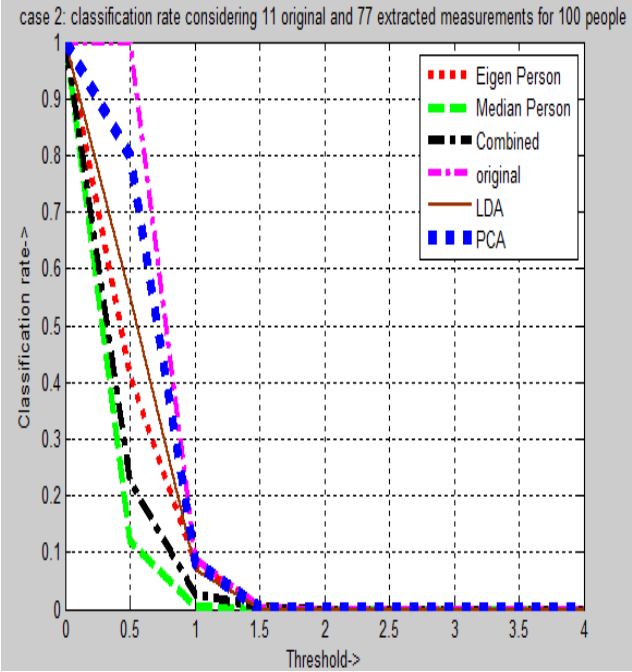


(a) 88 measurements and 100 people

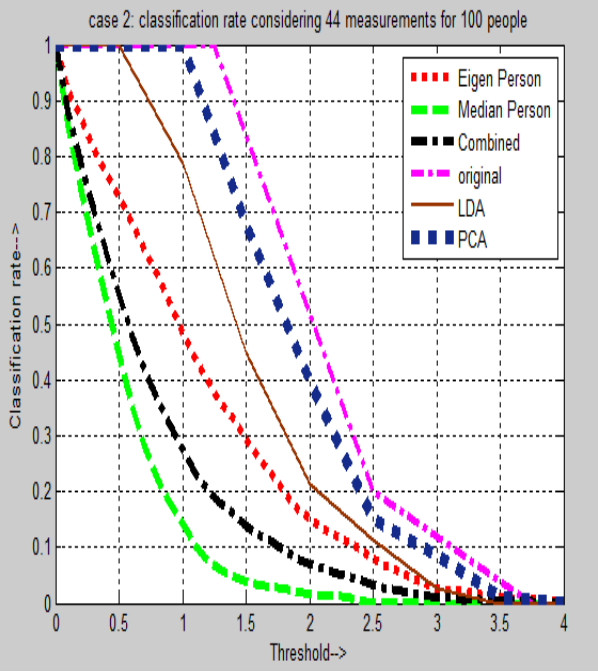


(b) 44 measurements and 100 people

Case 2: original person against all original people and their variations

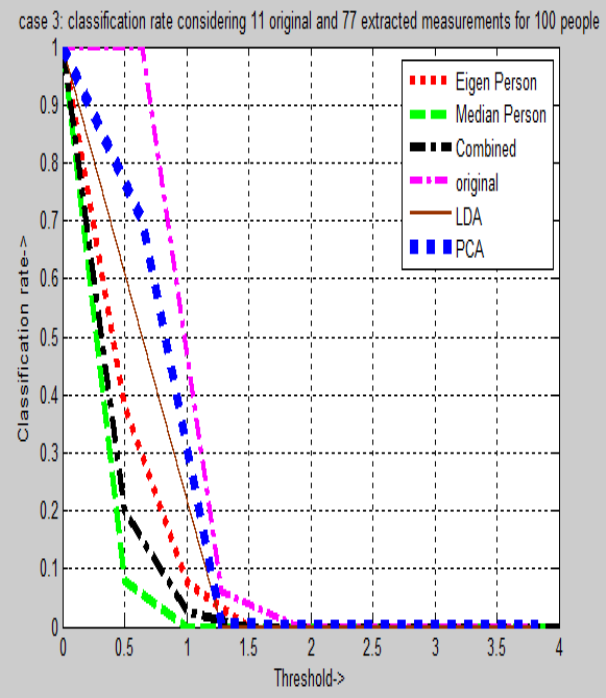


(c) 88 measurements and 100 people

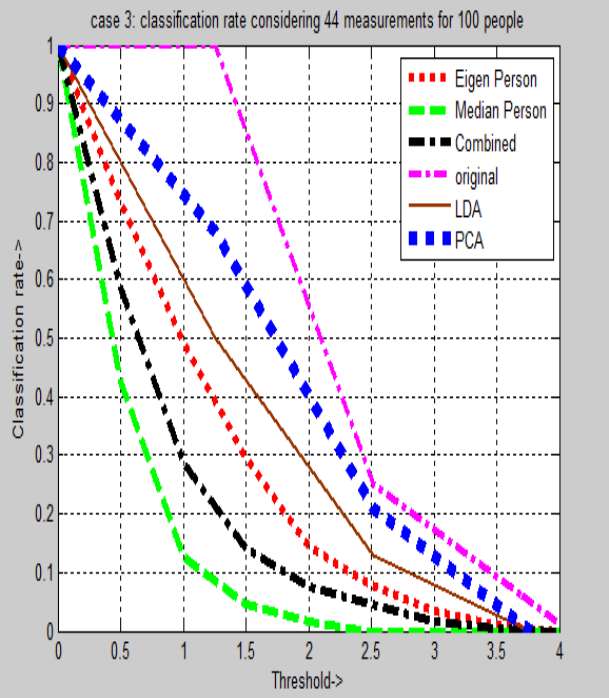


(d) 44 measurements and 100 people

Case 3: original person against their variations and remaining original people and their variations



(e) 88 measurements and 100 people



(f) 44 measurements and 100 people

Figure 11: Classification rate as a function of threshold for each case

At a particular threshold and given dataset, we can know the classification rate for different measurements in each case from the above plots. As can be observed in the above figures, the classification rate is better for SET-44 compared to 88 features (SET-11 original and 77 extracted features). Also as number of persons increase, the classification rate decreases.

5.5.3.1 Factor of age and gender

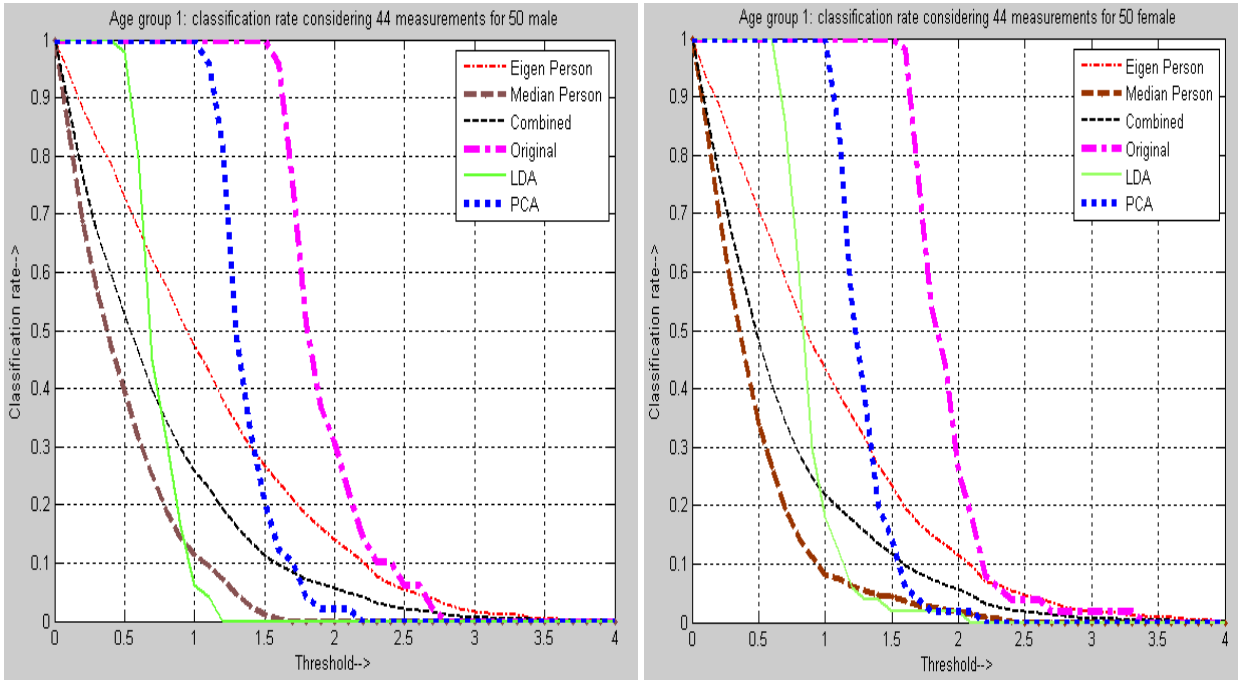
The classification plots in above figures are conditional distributions because we considered both male and female genders as a single group. We now consider each gender separately and look at their performance measure. We also consider the factor of age in distinguishing between humans. From the provided dataset, we divide the persons into following groups based on both age and gender:

(i) Age group 1 – (18 yrs to 25 yrs), (ii) Age group 2 – (26 yrs to 40 yrs),

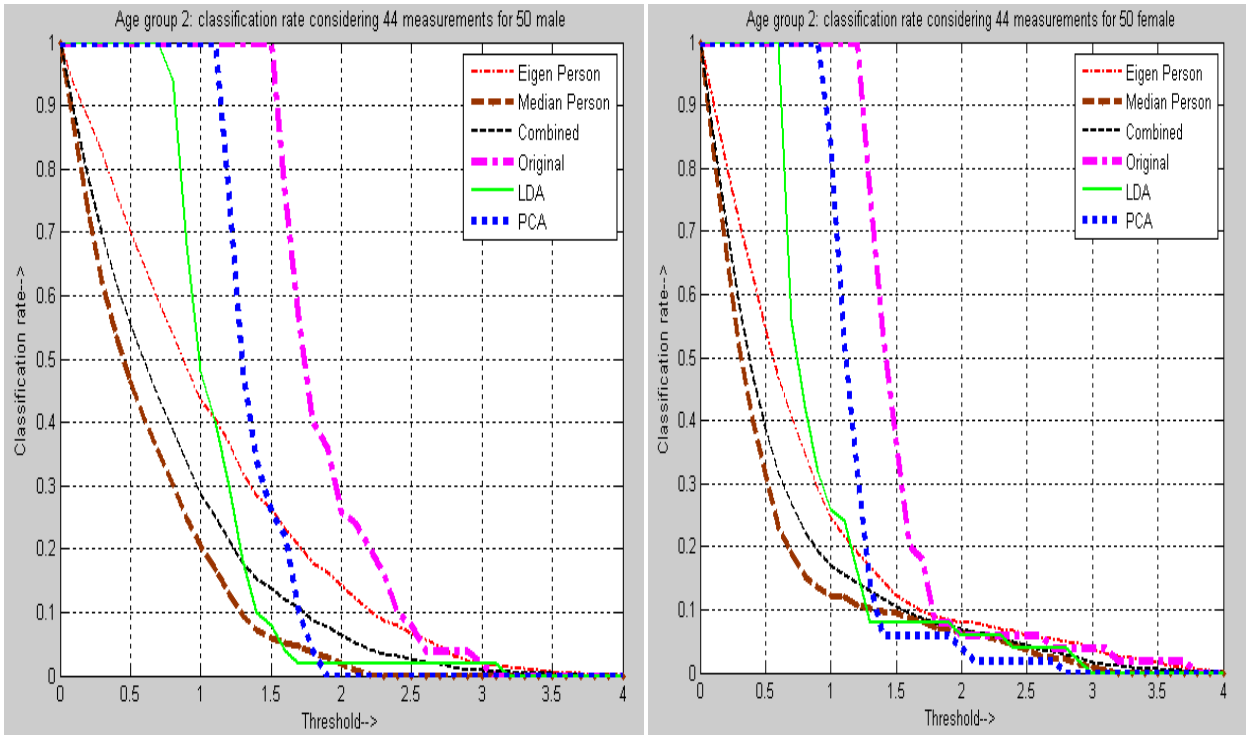
(iii) Age group 3 – (41 yrs to 59 yrs), and (iv) Age group 4 – (above 60 yrs)

The maximum age in Male group and Female group was 79 years and 69 years respectively. We plot the classification rates for each age group ((i) - (iv)) and gender (male or female). We

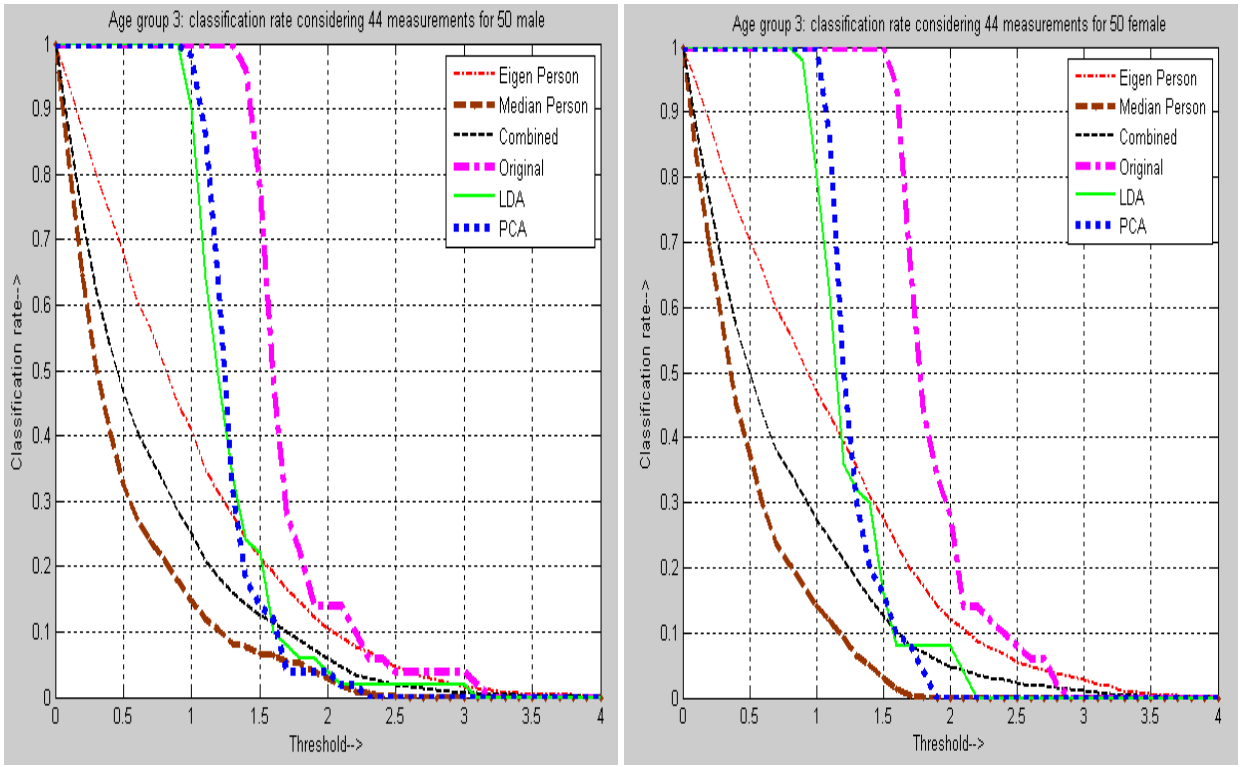
consider a training set of 50 male and 50 female in each age group. The results are shown in Figure 12. The male plots are on the left side and female plots to the right.



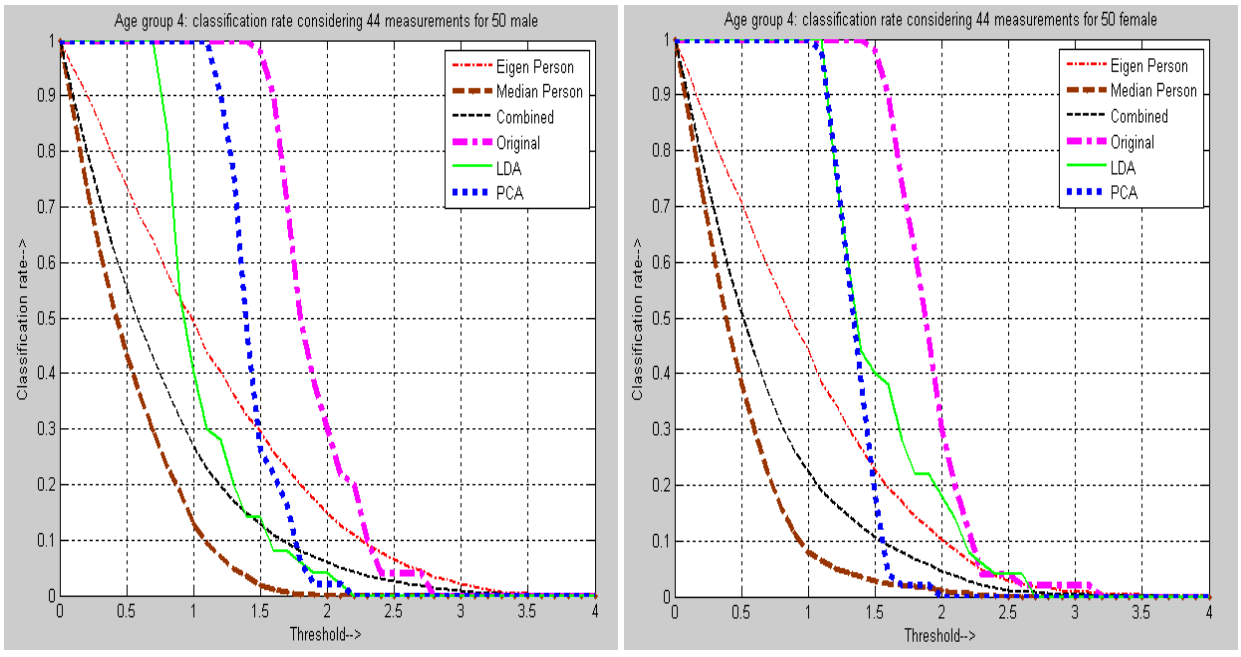
(a) Classification rate for 50 male and 50 female in age group (i) - (18 yrs to 25 yrs)



(b) Classification rate for 50 male and 50 female in age group (ii) - (26 yrs to 40 yrs)



(c) Classification rate for 50 male and 50 female in age group (iii) - (41 yrs to 59 yrs)



(d) Classification rate for 50 male and 50 female in age group (iv) - (above 60 yrs)

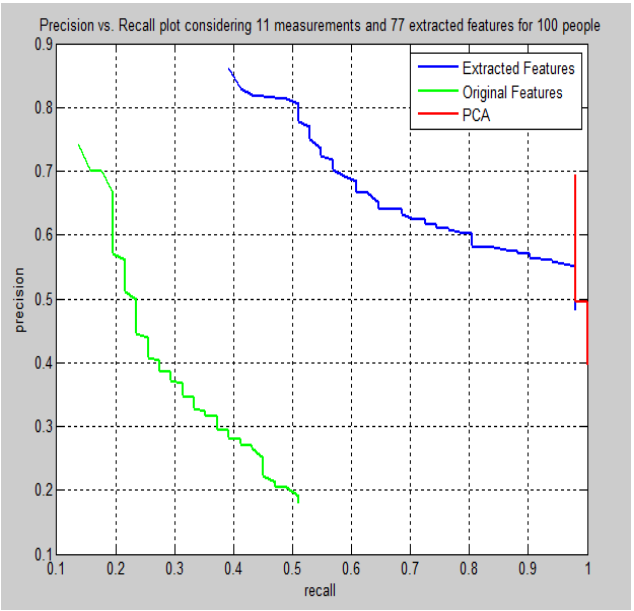
Figure 12: Classification rates for each age group considering male and female genders

The above plots give an idea of effect of age in determining the classification rate for the chosen training set of 50 male and 50 female. A closer look at the plots suggest that for age groups of 26 to 59 years, in this training set, the classification rate for female group is higher compared to male group. For the remaining age groups, the classification rates are close for both genders.

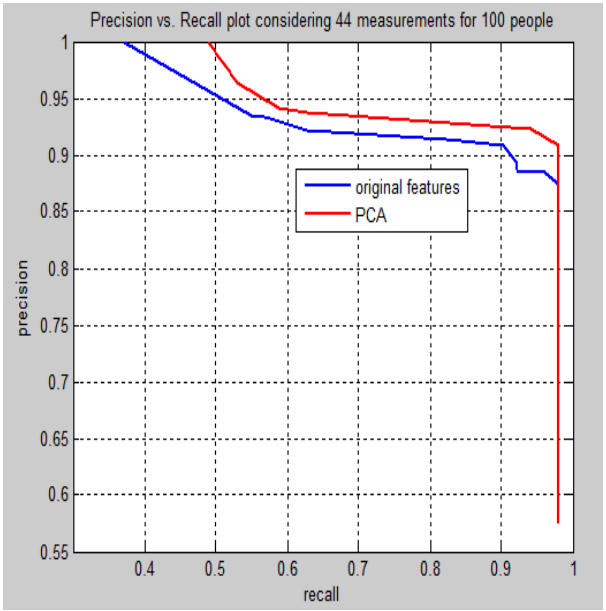
The individual classification rate of each gender in a particular age group is less compared to classification rate of both genders combined as a single group. This is observed by comparing above figures with Figure 10, where the people were from all age groups.

5.5.3 Precision vs. Recall

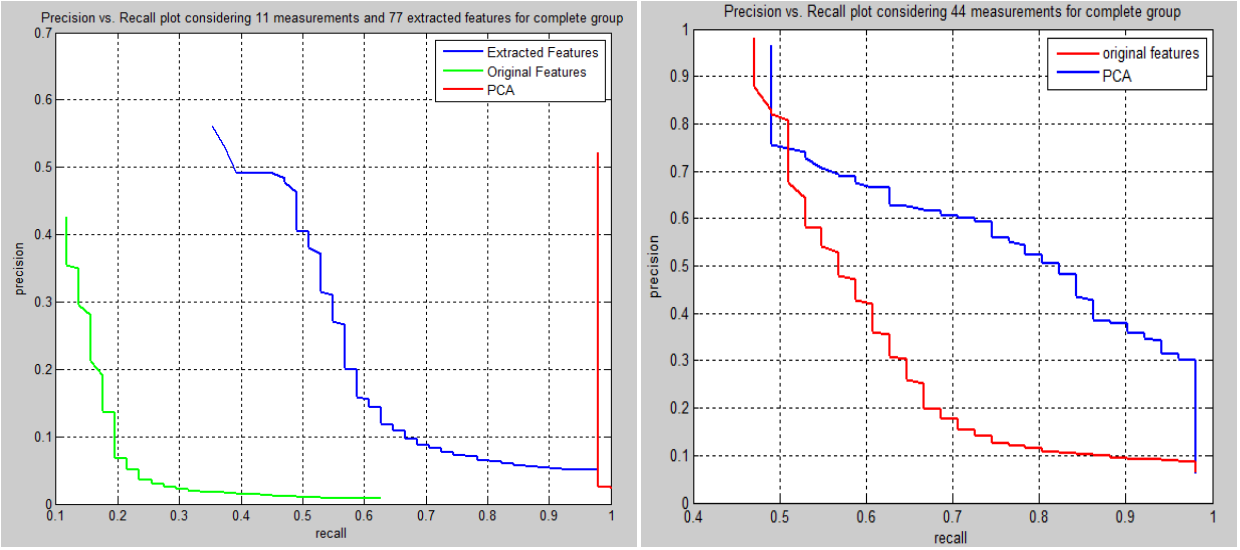
We compute precision and recall as a characteristic performance measure. As described in Equation (7), we calculate the precision and recall assuming N = 50 variations for each individual person, and a threshold of 0.1. The plots are shown in Figure 13. We show the plots for original and PCA features. The significance of using PCA is to discover if the selected measurements (SET-11 and SET-44) can be represented by a smaller number of measurements.



(a) 88 measurements and 100 people



(b) 44 measurements and 100 people



(c) 88 measurements over complete set

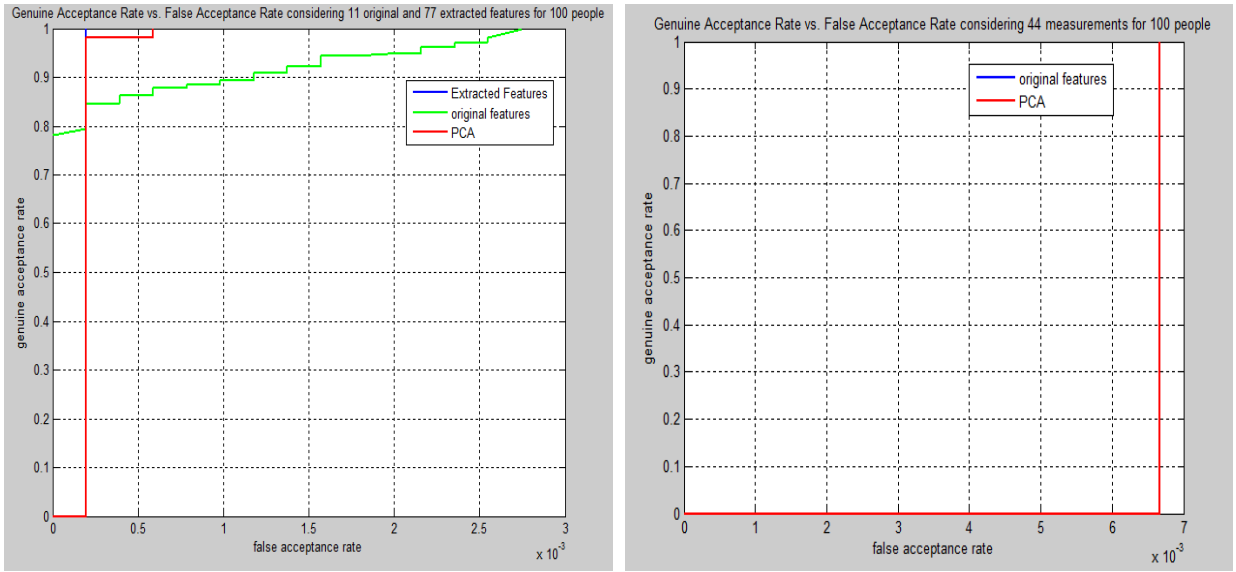
(d) 44 measurements over complete set

Figure 13: Precision vs. Recall curve for both training and complete set

Precision-recall curve represents accuracy of a system. From above results, it can be inferred that PCA is doing much better compared to original or extracted features. Also the accuracy is more for SET-44 compared to SET-11 features.

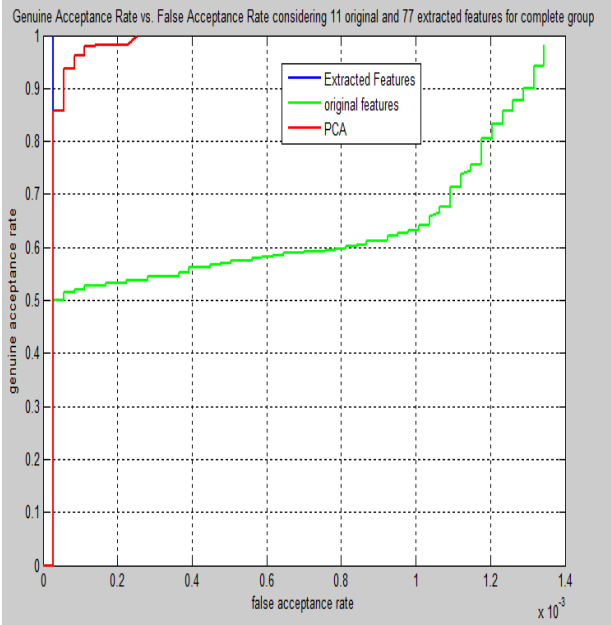
5.5.4 Genuine Acceptance Rate (GAR) vs. False Acceptance Rate (FAR)

Based on Equation (8), we calculate GAR and FAR, and plot them against each other for different measurements and sets. For this N=50 and threshold = 0.1. The plots are shown in Figure 14.

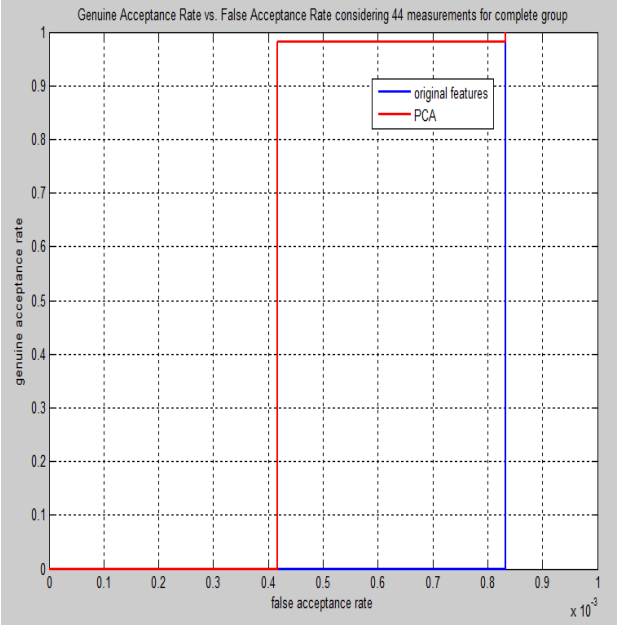


(a) 88 measurements and 100 people

(b) 44 measurements and 100 people



(c) 88 measurements over complete set

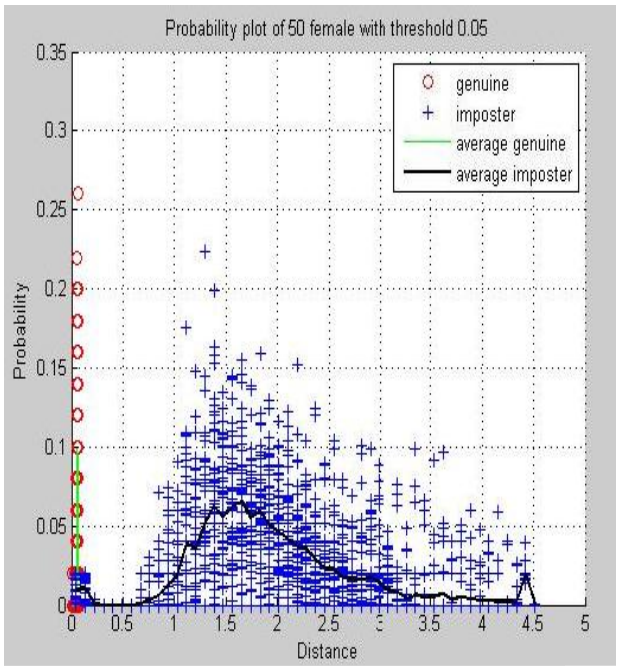


(d) 44 measurements over complete set

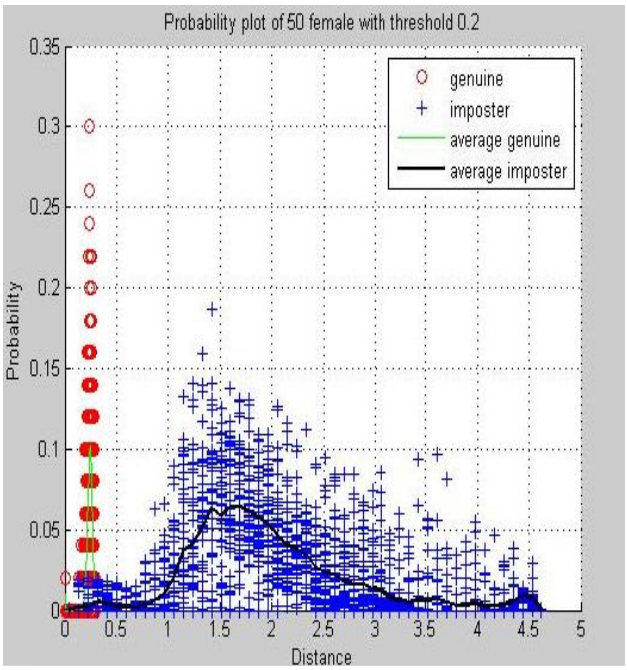
Figure 14: Genuine Acceptance Rate vs. False Acceptance Rate

The above results show performance based on gar and far. The performance is good when compared to other biometric systems and gives an initiative of performance of different feature sets used in the dataset.

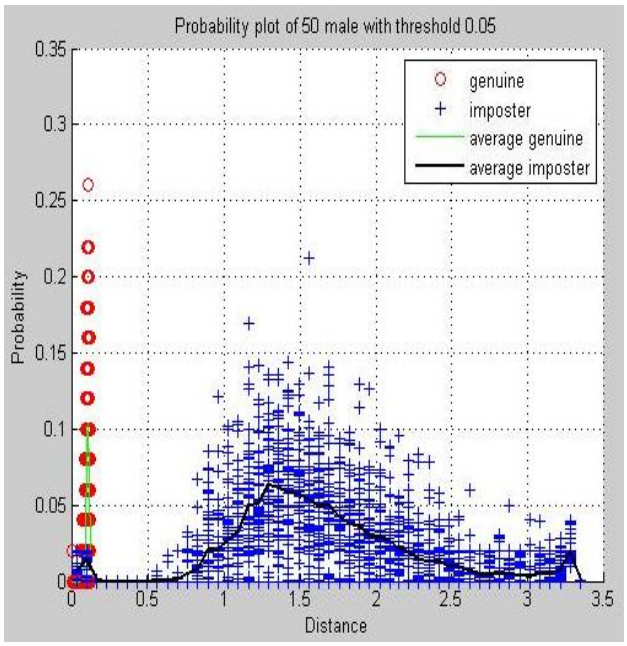
Later we plot detailed Genuine and Imposters probabilities to observe at what distance they are occurring. We introduce 50 variations of each person, where each measurement is varied by a threshold of 5% or 20%. The plots are shown in Figure 15 and Figure 16.



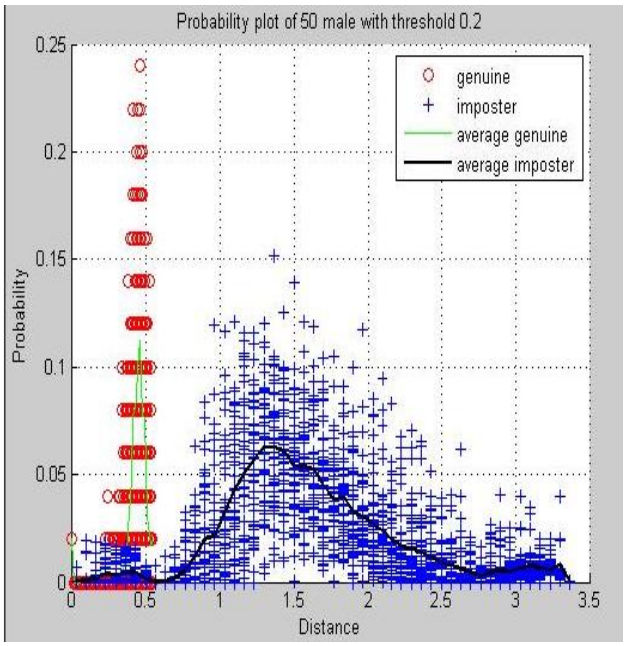
(a) 50 Female with 44 measurements and threshold 0.05



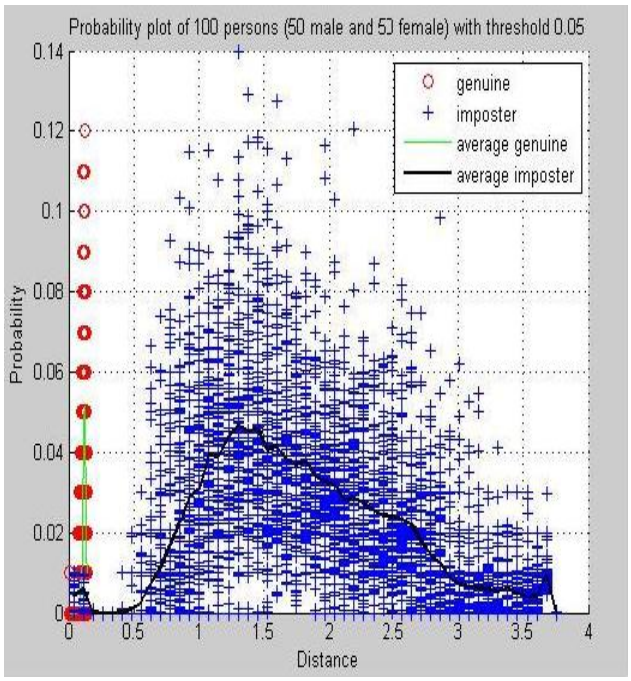
(b) 50 Female with 44 measurements and threshold 0.2



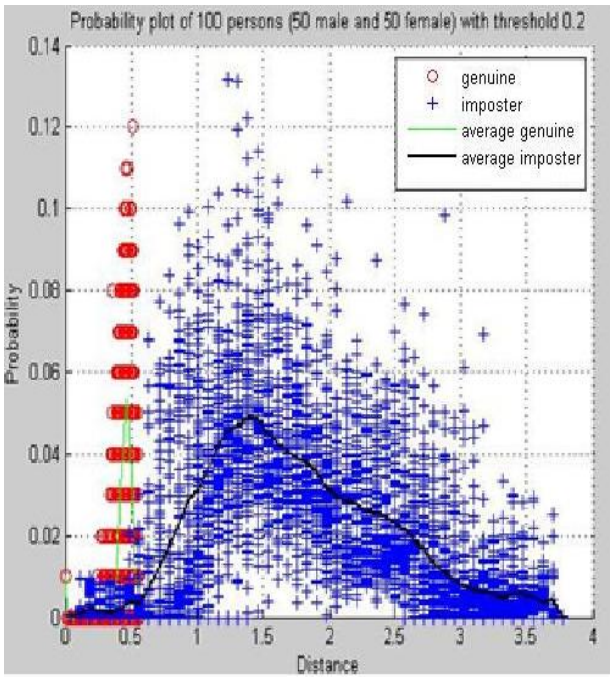
(c) 50 Male with 44 measurements and threshold 0.05



(d) 50 Male with 44 measurements and threshold 0.2



(e) 50 Male and 50 Female with 44 measurements and threshold 0.05

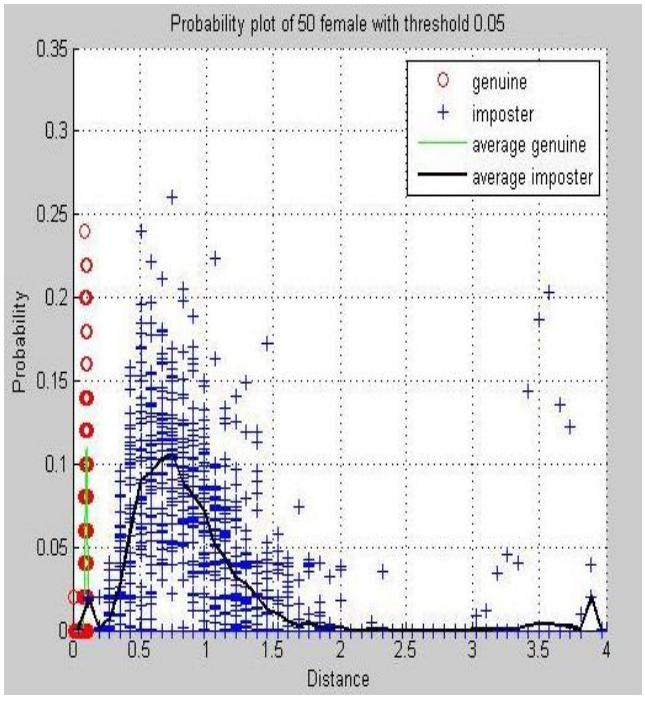


(f) 50 Male and 50 Female with 44 measurements and threshold 0.2

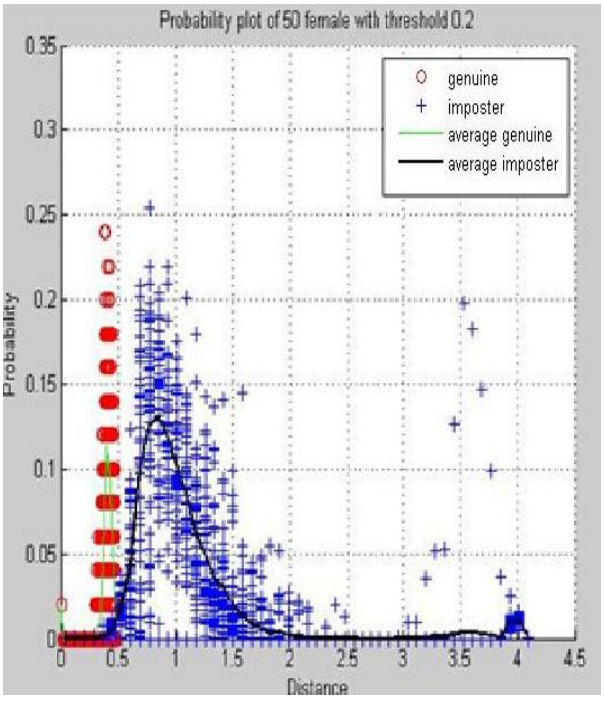
Figure 15: Probability of occurrence of Genuine and Imposters considering 44 measurements

.

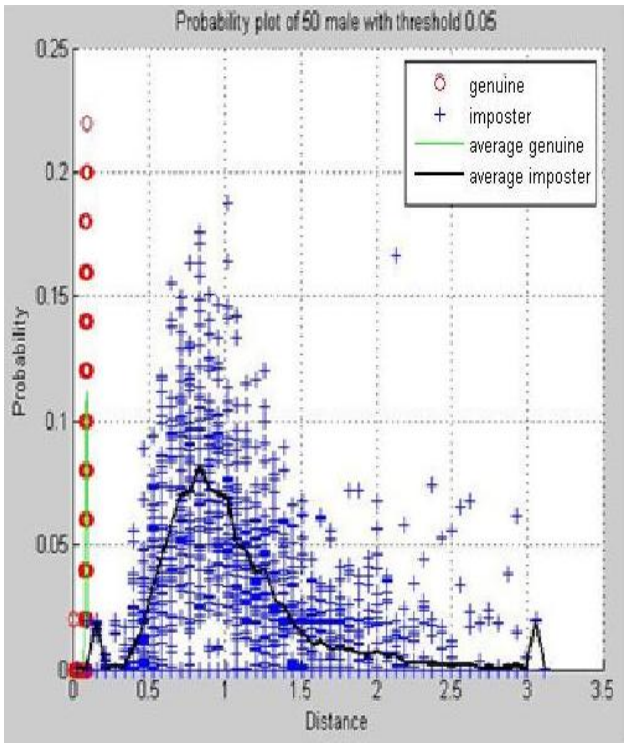
We similarly get the scatter plots for 88 measurements in Figure 16.



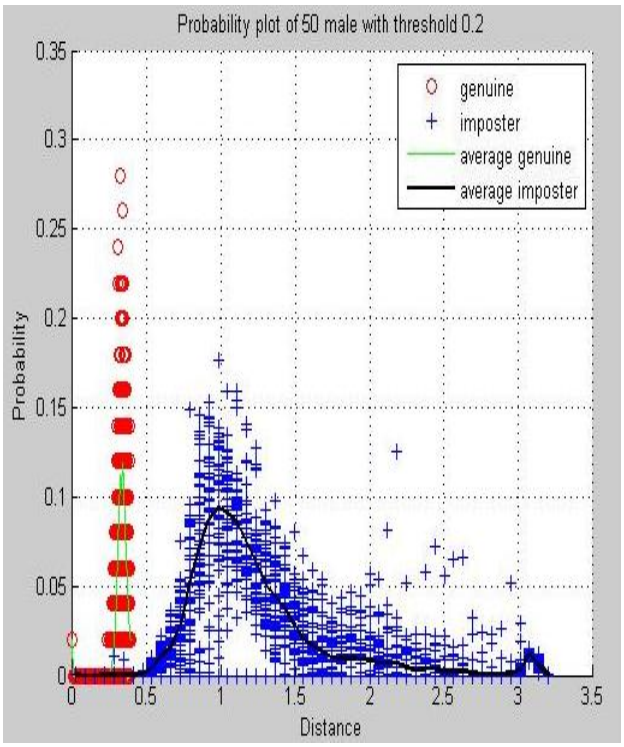
(a) 50 Female with 88 measurements and threshold 0.05



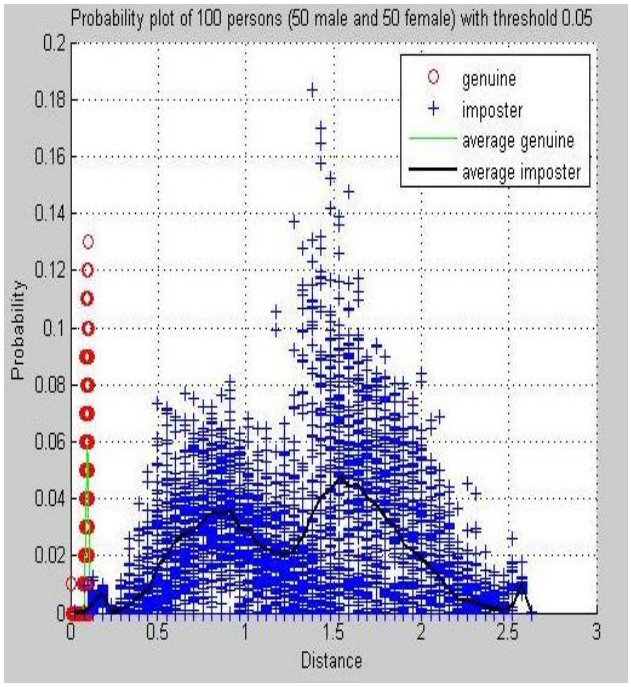
(b) 50 Female with 88 measurements and threshold 0.2



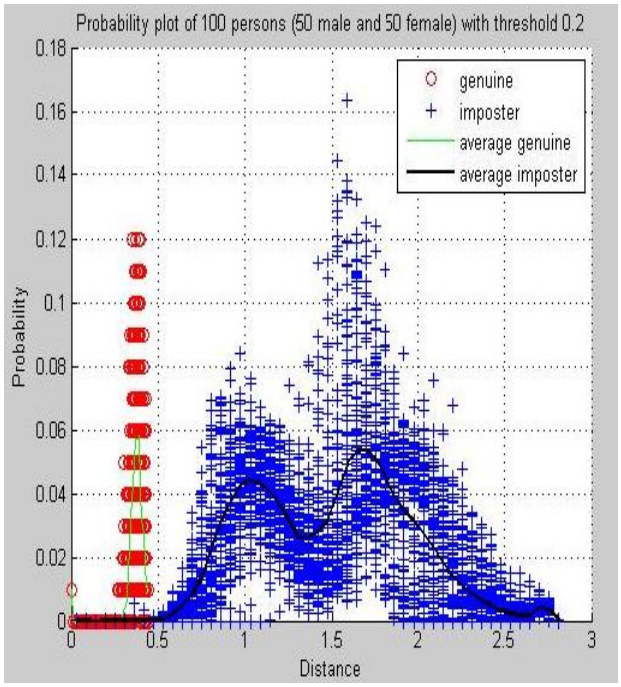
(c) 50 Male with 88 measurements and threshold 0.05



(d) 50 Male with 88 measurements and threshold 0.2



(e) 50 Male and 50 Female with 88 measurements and threshold 0.05



(f) 50 Male and 50 Female with 88 measurements and threshold 0.2

Figure 16: Probability of occurrence of Genuine and Imposters considering 88 measurements

The above results are very helpful in classifying a subject as genuine or imposter person. For an unknown person or for a variation of known person, we can classify him/her as an imposter or genuine person based on where their distance falls from other people in the dataset. Also given the dataset and threshold, we can set the cutoff point which marks off genuine and imposters to the best level.

The distribution of data in the above curves is not normal. This can be observed in Figure 17, which shows the hypothesis test for normality. The normal probability plot is created with "normplot" function in MATLAB. It combines normal probability plots with hypothesis tests for normality. The plus signs plot the empirical probability versus the data value for each point in the data. The y-axis values are probabilities from zero to one, but the scale is not linear. In a normal probability plot, if all the data points fall near the line, an assumption of normality is reasonable. Otherwise, the points will curve away from the line, as shown in this figure.

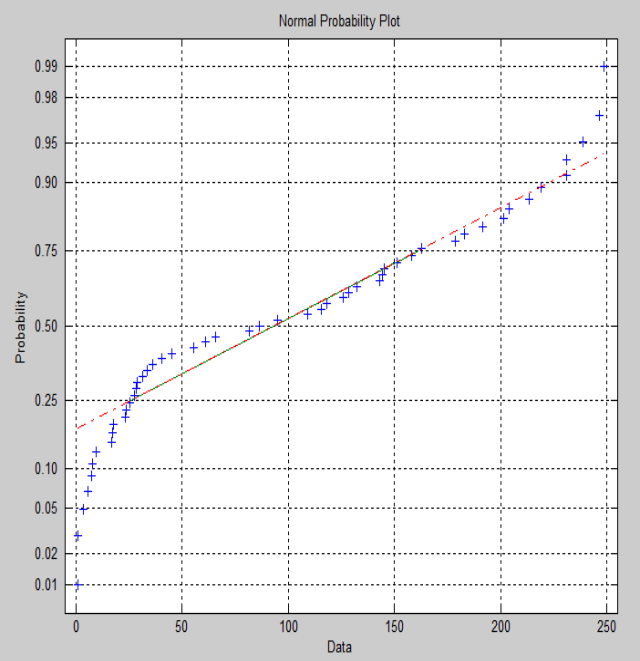
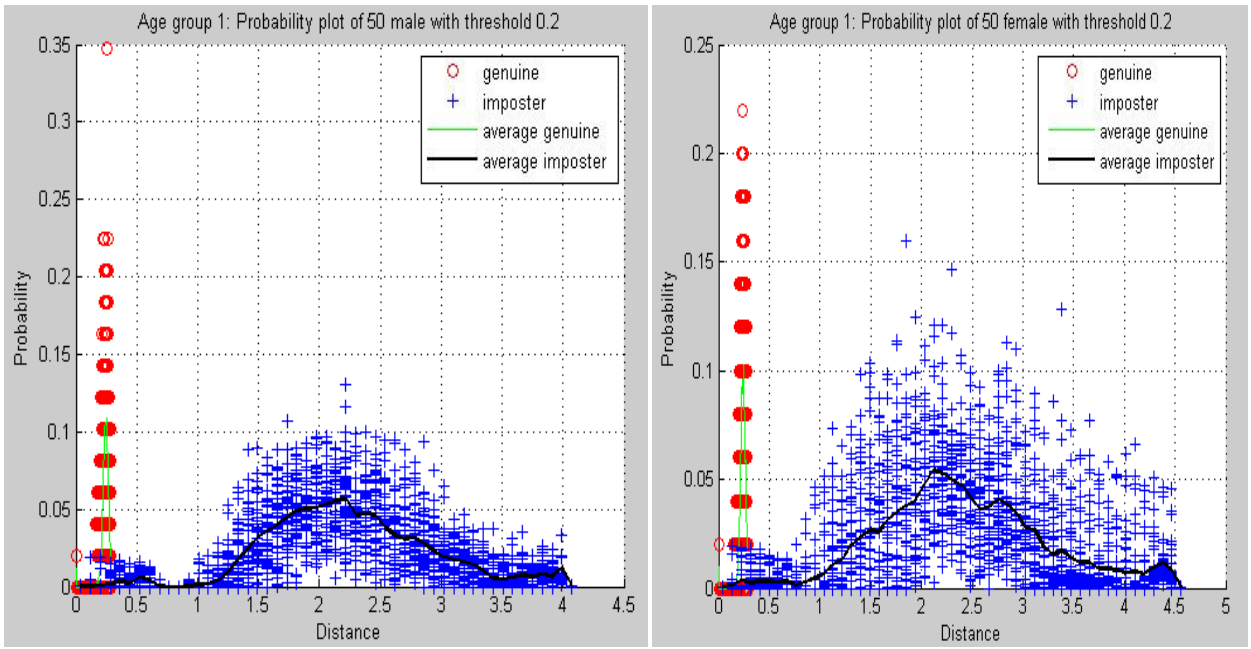


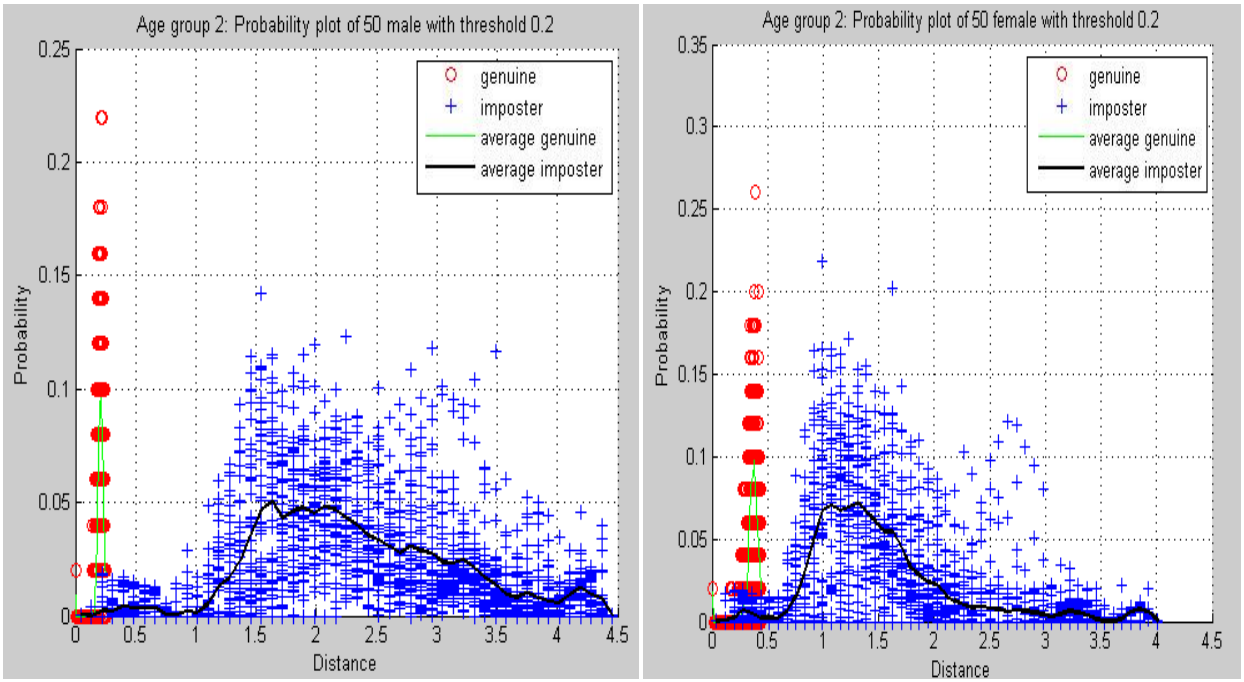
Figure 17: Normal probability plot of 50 male, 50 female and threshold 0.2

5.5.4.1 Factor of age and gender

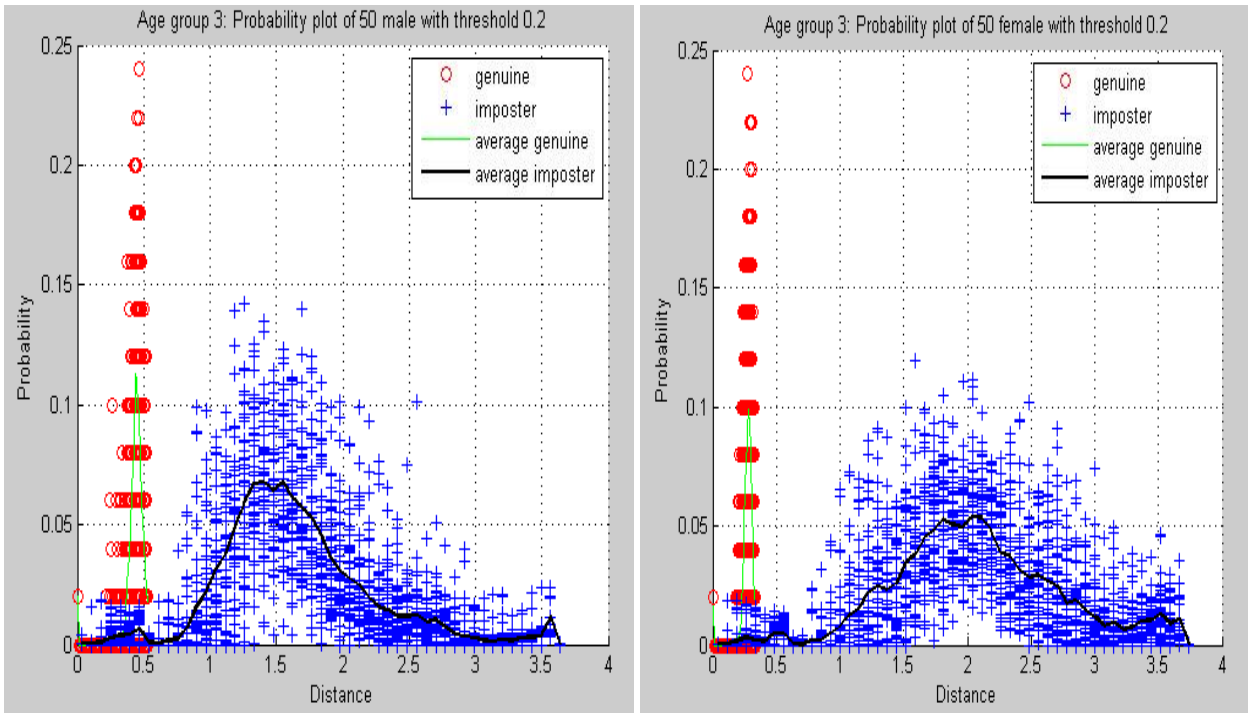
We also plot the frequency-distribution plots related to age and gender as discussed in Section 5.5.2.1. The different age groups are: (i) Age group 1 – (18 yrs to 25 yrs), (ii) Age group 2 – (26 yrs to 40 yrs), (iii) Age group 3 – (41 yrs to 59 yrs), and (iv) Age group 4 – (above 60 yrs). We also consider male and female genders in each age group. The results are shown in Figure 18 which shows the genuine-imposter probability curves.



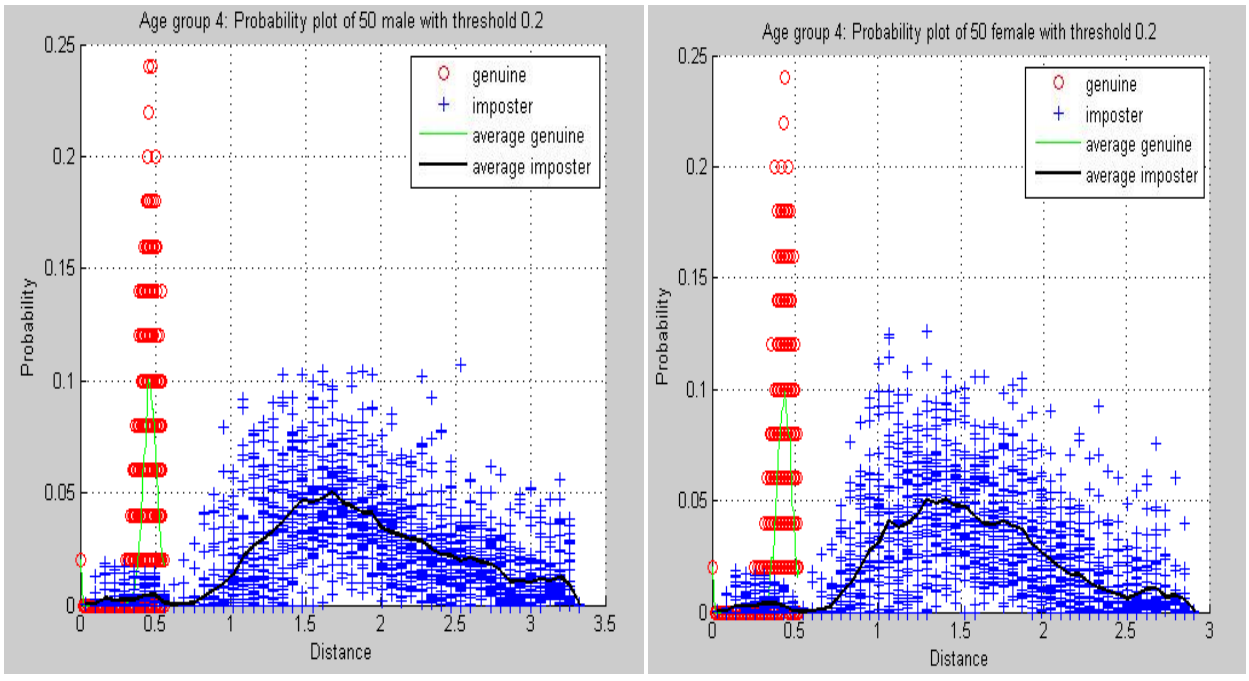
(a) Age group (i) - (18 yrs to 25 yrs)



(b) Age group (ii) - (26 yrs to 40 yrs)



(c) Age group (iii) - (41 yrs to 59 yrs)



(d) Age group (iv) - (above 60 yrs)

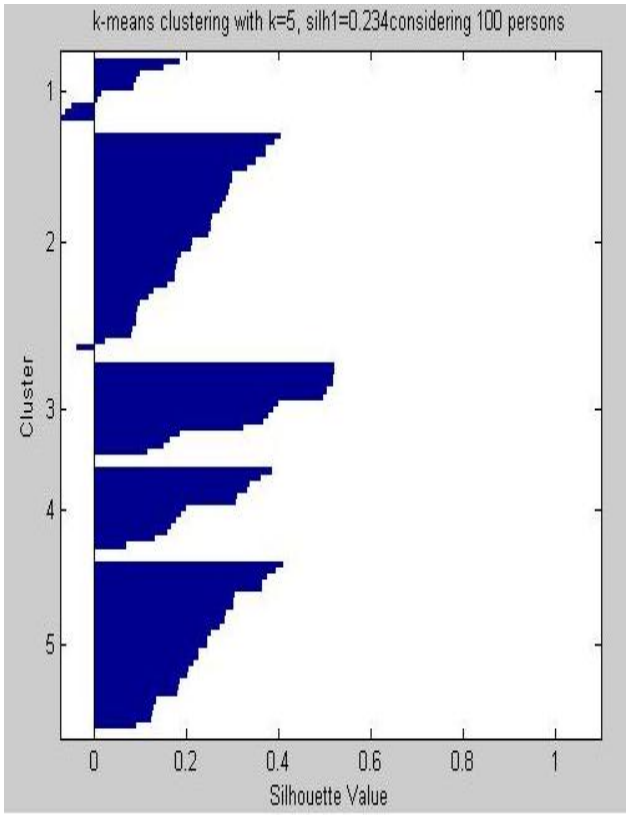
Figure 18: Probability of occurrence of Genuine and Imposters in each age group considering male and female genders

From these plots, we can decide the threshold at which a genuine person occurs in a group of imposters in each age group. Probability of occurrence of an imposter is more in female gender compared to male gender. Also genuine and imposter frequency is more in combined genders when compared to genders as a single group.

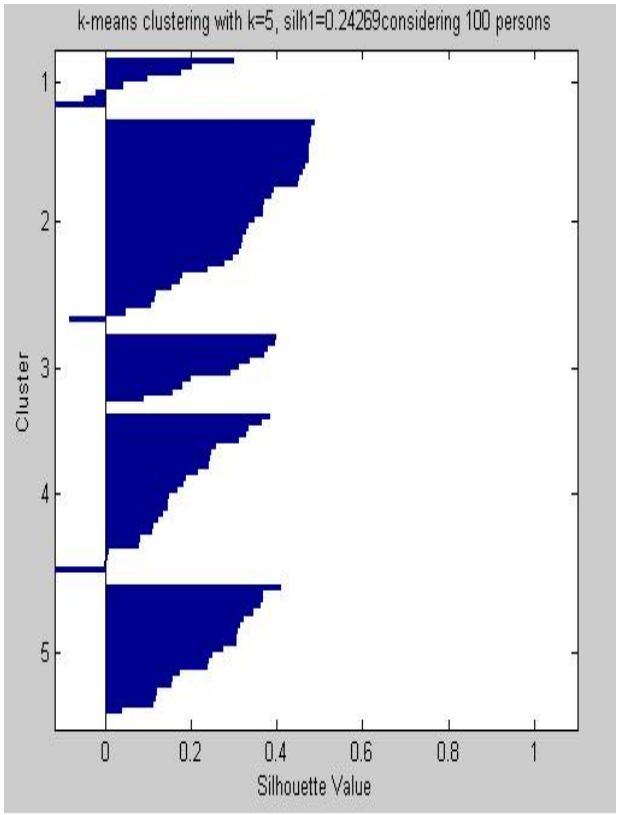
In Figure 16 (e) and (f), we get a bimodal graph. This can be reasoned to having both male and female genders as single set in computing the classification rates, rather than as individual groups.

5.5.5 Clustering

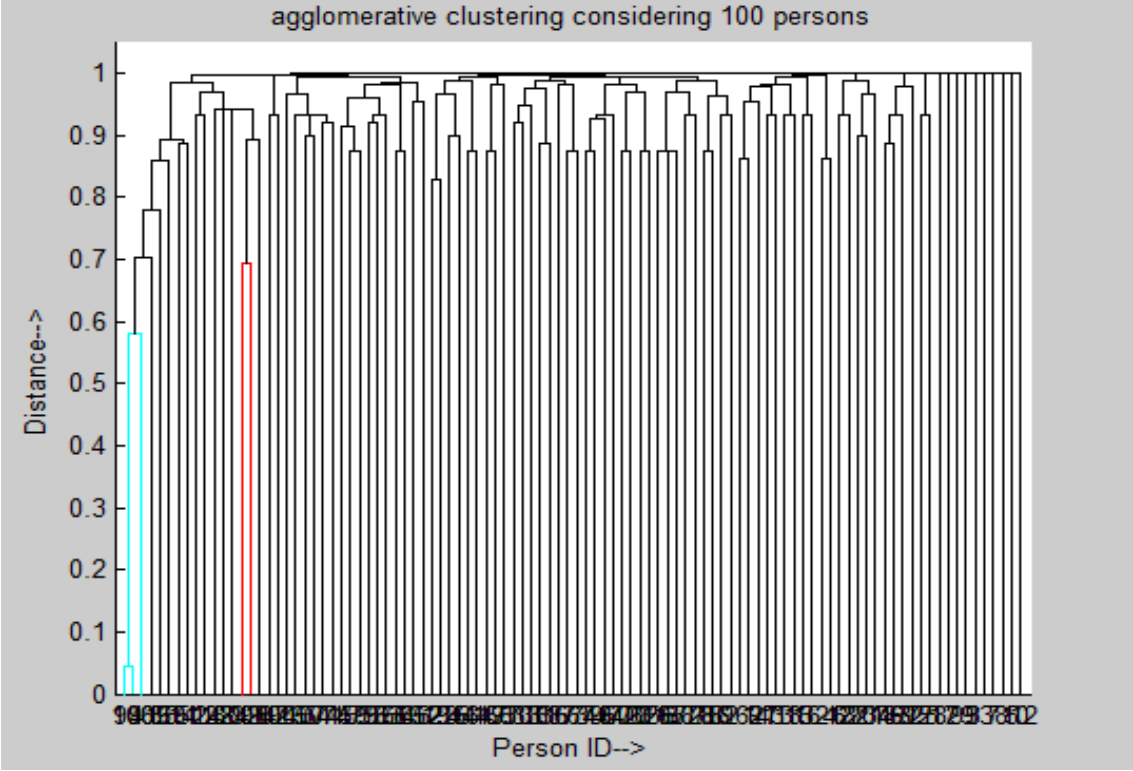
Using *K*-means and Agglomerative clustering, we get the clusters for 44 and 88 measurements for a training set of 100 people. They are shown in Figure 19.



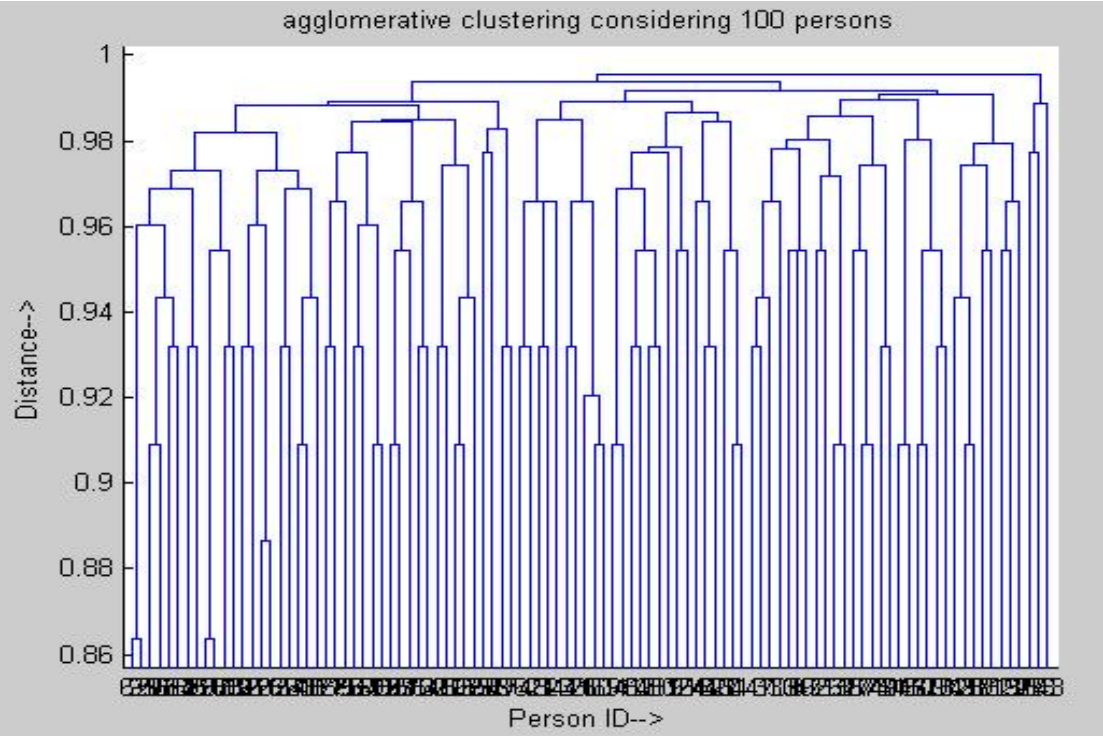
(a) *K*-means clustering of 100 people considering 88 measurements



(b) *K*-means clustering of 100 people considering 44 measurements



(c) Agglomerative clustering of 100 people considering 88 measurements



(d) Agglomerative clustering of 100 people considering 44 measurements

Figure 19: Clustering of a training set of 100 people considering 44 and 88 measurements

From the above results, we can observe that Agglomerative clustering is more informative compared to *K*-means clustering. In Figure 19(c) and (d), if we take some threshold distance of, say, 0.98, we can know how many persons are clustered above that threshold by marking a cut-off line. All persons above that line are clustered at that threshold.

5.6 Discussion

From the above results, we get to know the performance of CAESAR dataset. We initially selected 11 measurements (SET-11), and later did feature extraction on these measurements as mentioned in Section 4.5. We later consider 44 measurements (SET-44), and impart 50 variations on each of the original persons in both these sets while varying a threshold.
On each of these features, we carry out performance measures resulting in characteristic plots. From the initial figure of scatter plot, we get an idea about the relation between selected measurements. Later we generate the histogram over the complete dataset to know the distribution of these features over the entire population.

We later compute distance plots for both set of measurements. It can be observed that as the number of features considered increase from 11 to 44, the person-to-person distance increases. From Table 9 it can be detected that the minimum distance threshold increases from 0.5251 to 0.9116 for the complete set when the features increase from 11 to 44 respectively. Hence the persons with 44 features are better separated than the persons considered with just 11 features.

Moving further to the classification rate plots, we defined classification rule as number of dissimilar persons whose Euclidean distance is greater than some threshold (Equation (5)). Based on this, we plotted the classification plots for 44 features and 88 features. The plots of classification rate of eigen Person, median Person, and average Person for both male and female groups are shown in each of the varying cases 1, 2 and 3. At any given threshold, the classification rate is better for women than for men. From the results, we can observe that euclidean distance yields the higher classification rate compared to other distances. Also considering 44 features gives a better classification rate compared to 11 measurements and its extracted features. This can be attributed to the fact that the 11 features selected initially are not reliable in identifying people. Also we can say that metrology can distinguish people, by taking some threshold and below which if it finds just one person.

Hence distance plots and classification rate plots are helpful in distinguishing between persons.

From the precision-recall plots, we note that practically the system is able to retrieve correctly the variations of original persons. We get to know about the accuracy of measurement in identifying persons. It can be observed from the results that considering 44 measurements performs better than 88 features.

The genuine acceptance rate and false acceptance rate is a ROC curve showing the performance of the system in retrieving genuine persons even with variations. We can observe that ROC curve is good enough in detecting a varied person in the dataset. For the 11 measurements and its variations, we see that features extracted through PCA perform better compared to other extracted features or the 11 original features.

We later plotted the frequency plots for genuine and imposters which is practically important. From these plots, we can decide on the threshold that can be used in marking off between imposters and genuine persons. We can know the Type-I and Type-II errors which are critical in performance of a biometric system. At some distance, we come across some imposters acting as genuine persons. But we can identify them because they are within some distance threshold.

Next the cluster plots are shown, using *K*-means and agglomerative clustering. Agglomerative clustering shows the distance threshold at which a particular number of persons have been clustered. If we select a threshold at some distance, the portion above this line will convey useful information for future study.

Chapter 6: Conclusion

There are different recognition methods for human recognition. All the methods have their advantages and disadvantages. The major disadvantage of these recognition schemes is the human under study should be within the proximity of measuring instrument. But in practice, this may not be the case always. In such a situation we can make use of video metrology.

The main objective of this work has been to evaluate the performance of human metrology in distinguishing between persons using the CAESAR dataset. We can conclude that metrology can be used in distinguishing one person from other persons at a particular threshold. This can be observed from the distance plots, classification rate, precision-recall plots, ROC curves, and clustering methods.

Also feature extraction methods are used as described in this thesis. It can be said that PCA based features produce better results compared to other feature extraction methods or original measurements. This is observed from genuine acceptance rate and false acceptance rate plots.

Human metrology is fair enough in retrieving an original person from a pool of other original people and their variations. We can identify a genuine person from imposters.

Based on the results generated for training set and the complete set, we can say that human metrology is good in discriminating between persons, but care has to be taken to ensure that the parameters (like variation threshold, number of variations, or the distance threshold) are within some bounds.

From this thesis, we can say that Human metrology can be used in Biometrics to distinguish, identify and verify humans.

Future Directions

Here we discuss some future directions which can be carried out related to this work.

The performance measurement procedure discussed in this thesis can be used on real video to distinguish between humans. Also a larger training set can be employed and performance can be achieved.

This work can be extended to “allometric scaling”, where the objective is to generate synthetic variations of human models or animals based on their shapes and sizes [14].

References:

[1] Feng Guo and Rama Chellappa, “Video Metrology Using a Single Camera,” *IEEE Transactions on pattern analysis and Machine Intelligence*, Vol. 32, No. 7, July 2010 1329.

[2] http://store.sae.org/caesar/

[3] Jain, Anil K.; Flynn, Patrick; Ross, Arun A, *Handbook of Biometrics,* Springer 2006

[4] Electronic Privacy Information Center. "Face Recognition." January 19, 2006.

[5] Anil K. Jain, Arun Ross, and Salil Prabhakar, “An Introduction to Biometric Recognition.”

[6] Akita, K. 1984, “Image sequence analysis of real world human motion”, Pattern Recognition, 17(1):73-83, 1984

[7] Josef Bigün, Gérard Chollet, Gunilla Borgefors, “Book Audio-and video-based biometric person authentication,” *First International conference*, AVBPA '97.

[8] Brett Allen, Brian Curless, Zoran Popovic, “Exploring the space of human body shapes: Data-driven synthesis under anthropometric control,” *Proceedings of the SAE Digital Human Modeling for Design and Engineering Conference,* Rochester, MI. June 15-17, 2004.

[9] Brett Allen, Brian Curless, Zoran Popovic, “The Space of human body shapes: reconstruction and parameterization from range scans,” *ACM Transactions on Graphics (ACM SIGGRAPH 2003), 22, 3,* 587-594.2003.

[10] Afzal Godil, Patrick Grother, Sandy Ressler “Human Identification from Body Shape,” pp.386, *Fourth International Conference on 3-D Digital Imaging and Modeling (3DIM ’03)*, 2003.

[11] United Stated Air Force Research Laboratory, *Civilian American and European Surface Anthropometry Resource (CAESAR),* Final Report Volume II: Demographic and Measurement Descriptions, 2002.

[12] R. Lymann Ott, Michael Longnecker, *An Introduction to Statistical Methods and Data Analysis*, 6th Ed., Cengage Learning, 2010.

[13] Don Adjeroh, Deng Cao, Marco Piccirilli, and Arun Ross, "*Predictability and Correlation in Human Metrology".*

[14] T. McGraw, T. Kawai, J. Richards, "*Allometric Scaling for Character Design",* Computer Graphics Forum, 2010.

[15] Afzal Godil, Sandy Ressler "Retrieval and clustering from a 3D Human Database based on Body and Head Shape," *in proceedings of 2006 Digital Human Modeling for Designand Engineering,* July 2006, Lyon France.

[16] Christopher M. Bishop, *Pattern Recognition and Machine Learning*, Springer 2006.

[17] P. Jonathon Phillips, Patrick Grother, and Ross Micheals "Evaluation Methods in Face Recognition," *Handbook of face recognition*, Pages 328-348, 2005.

[18] Chorkin Chan and Pak-bong Wong, "A Branch and Bound Decision Tree Bayes Classifier for Robust Multi-font Printed Chinese Character Recognition," 11th - 13th November, 1992 Melbourne, Australia.

[19] C. Madden, M. Piccardi, "Height Measurement as a session-based Biometric for people matching across disjoint camera views," *Image and vision computing New Zealand,* pp. 282-286, 2005.

[20] D. Decarlo, D. Metaxas, M. Stone "An Anthropometric face model using variation techniques," *Proceedings of ACM SIGGRAPH 98, ACM Press,* Computer Graphics Proceedings, Annual Conference Series, 67-74. 1998.

[21] Hye-won Seo and Nadia M. Thalmann, "An automatic modeling of human bodies from sizing parameters" *Proceedings of the 2003 Symposium on Interactive 3D Graphics,* ACM Press, 19-26. 2003.

[22] S. I. R. Ellison, T. Fearn, “Characterizing the performance of qualitative analytical methods: Statistics and terminology,” *TrAC, Trends in analytical chemistry* vol. 24, pp. 468-476, 2005.

[23] Peter Somol, Pavel Pudil, Josef Kittler, “Fast Branch & Bound Algorithms for Optimal Feature Selection*,” IEEE Transactions on Pattern Analysis and Machine Intelligence,* vol. 26, no. 7, pp. 900-912, July 2004.

[24] Richard O. Duda, Peter E. Hart, David G. Stork, *Pattern Classification*, 2nd Ed., Wiley Interscience, 2000.